Molecular basis of recombination, the importance of recombination in genetics

DNA recombination

- exchange of segments of DNA molecules **between chromosomes**
- often occurs during **meiosis** in sexual reproduction - replacing **parts of homologous chromosomes**
- increase of **genetic diversity in the offsprings** - an evolutionary advantage for offsprings
- It exists in **prokaryotic cells** (after transfer of foreign DNA by transformation, transduction or conjugation)

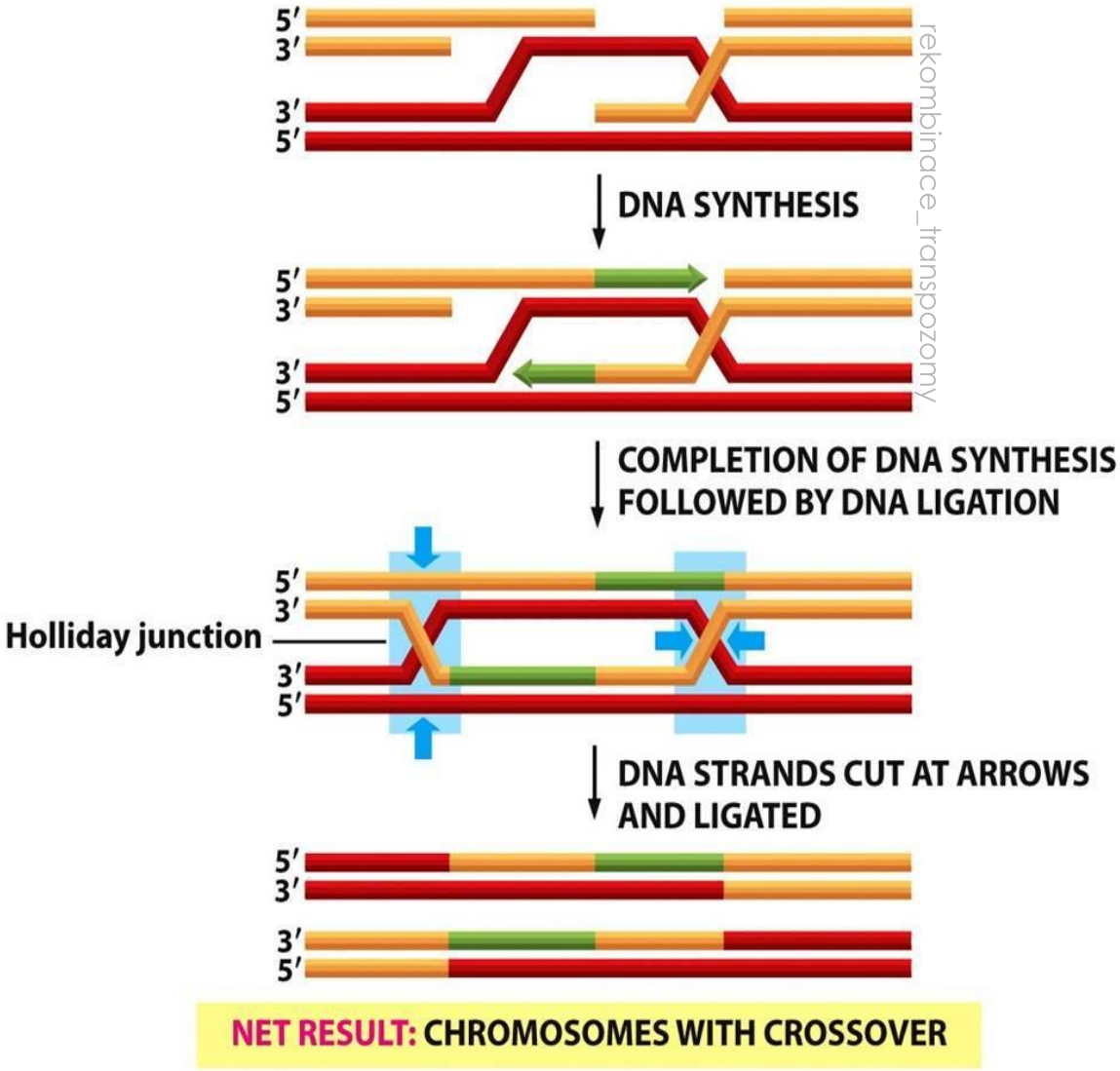
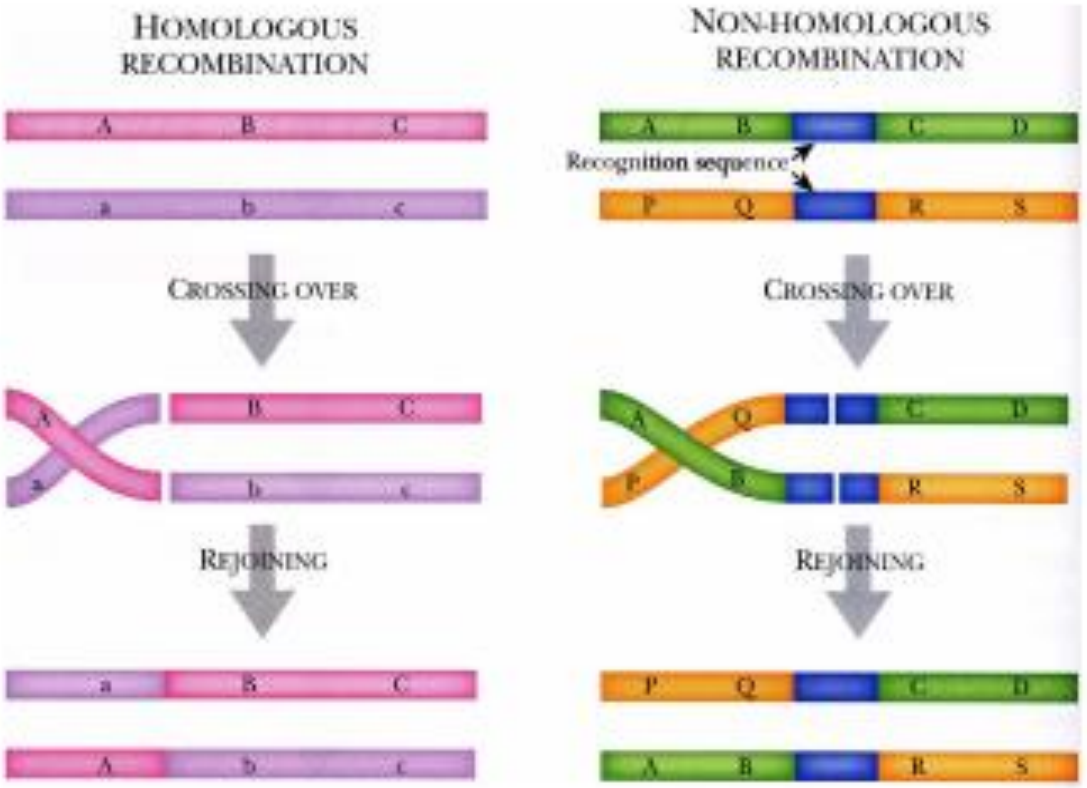


Figure 6-31 part 2 of 2 Essential Cell Biology 3/e (© Garland Science 2010)

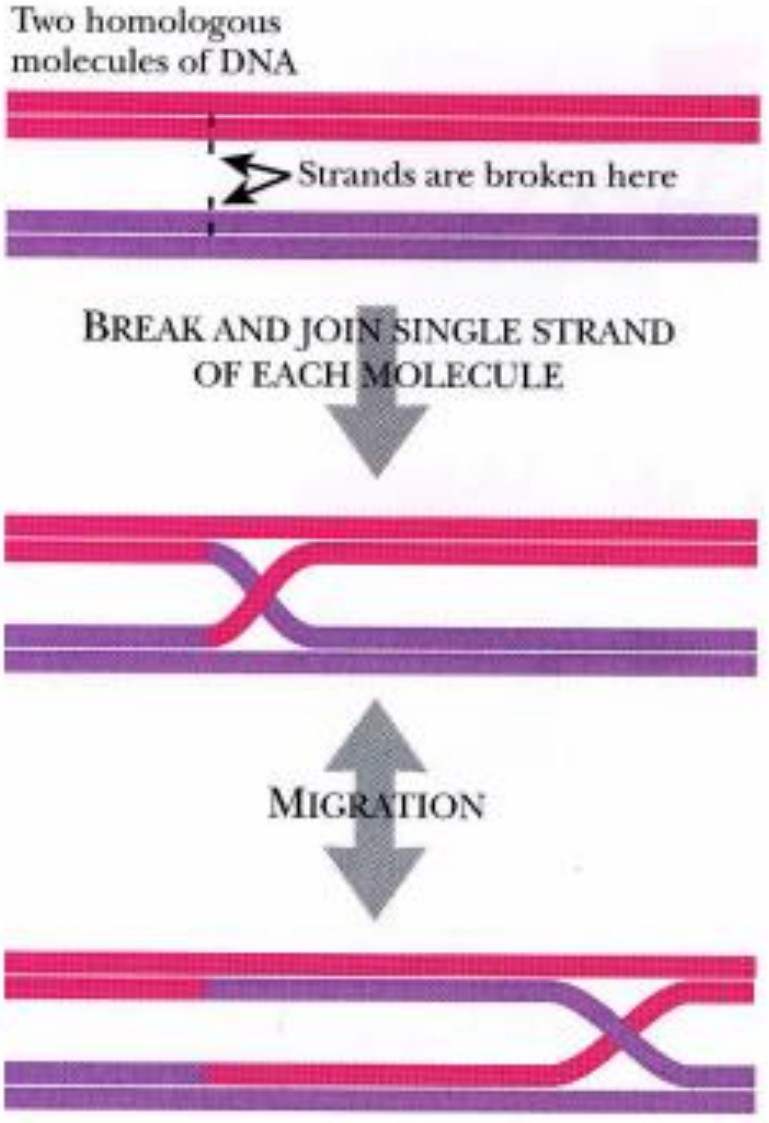
Homologous and non-homologous recombination

- **homologous recombination**
- allows the exchange of **genetic material** between chromosomes that are so similar that can lead to **base pairing** between them
- **common between the two copies of the same chromosome (in meiosis)**
- **non-homologous recombination**
- rarer, does not require sequence homology
- requires specific proteins

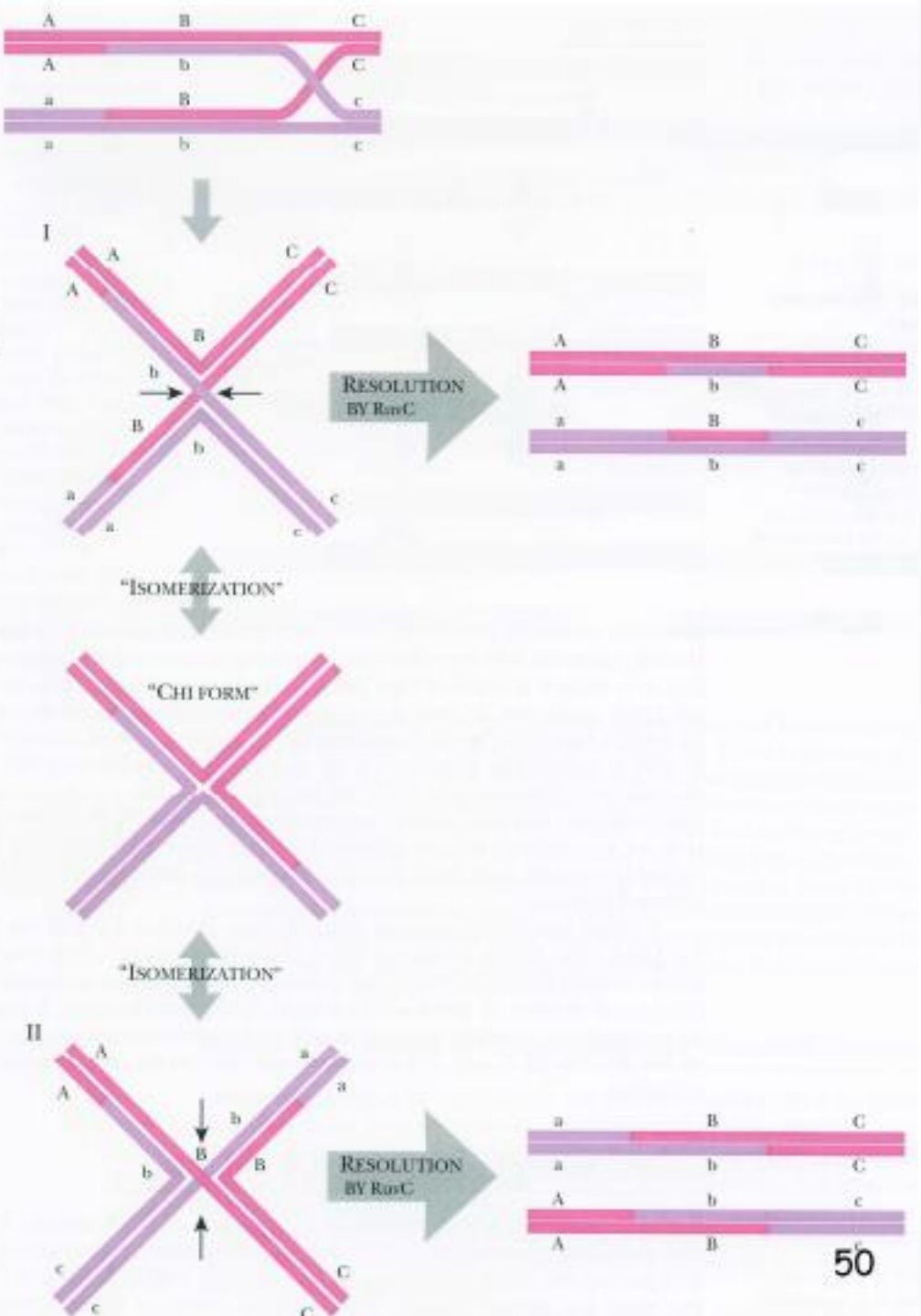


Molecular basis of homologous recombination

- reciprocal recognition of homologous segments of double-stranded DNA
- interruption of one strand of each helix
- replacement of strands
- reuniting to form **Holliday structure**

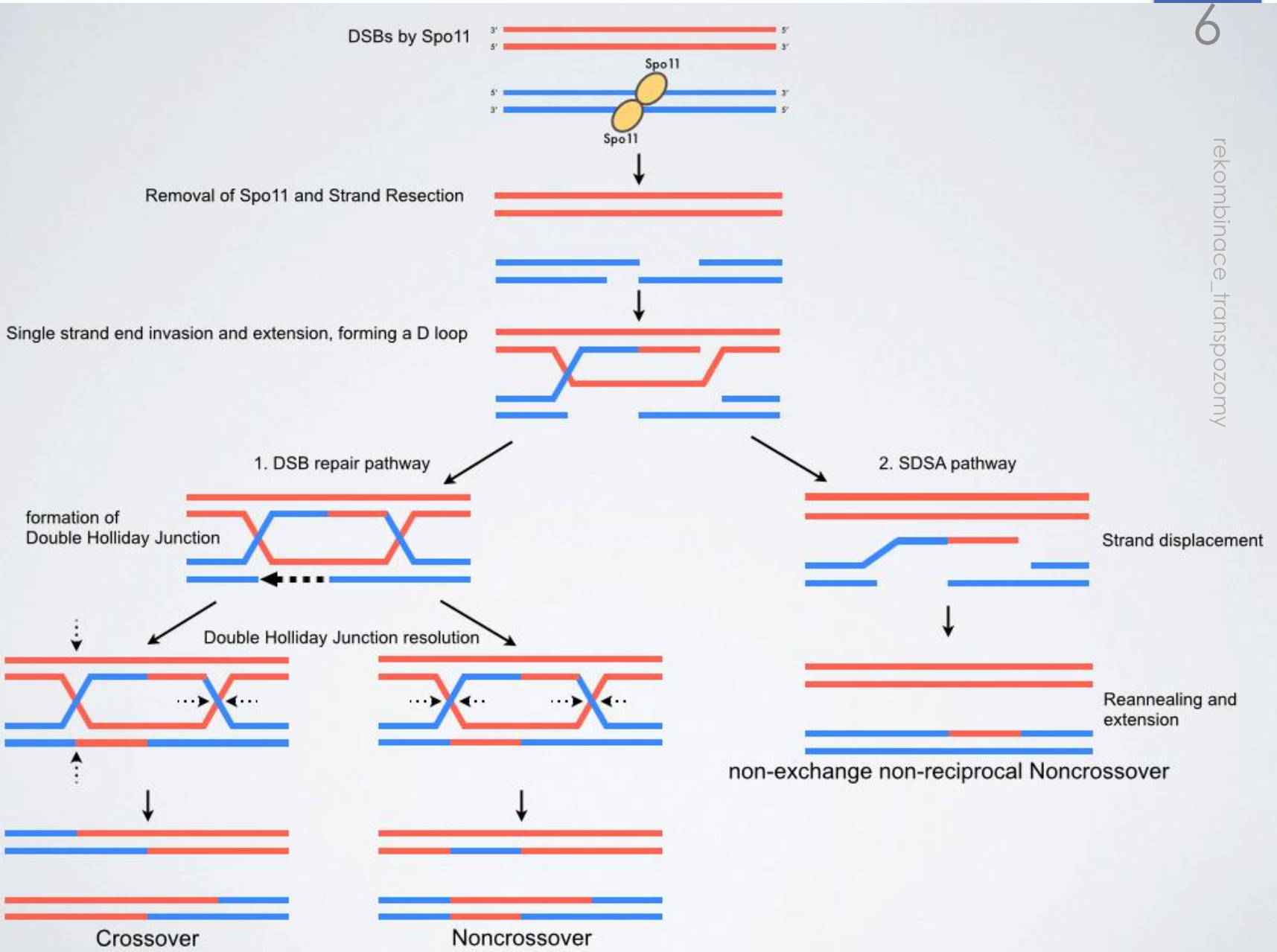


Holliday structure can isomerize



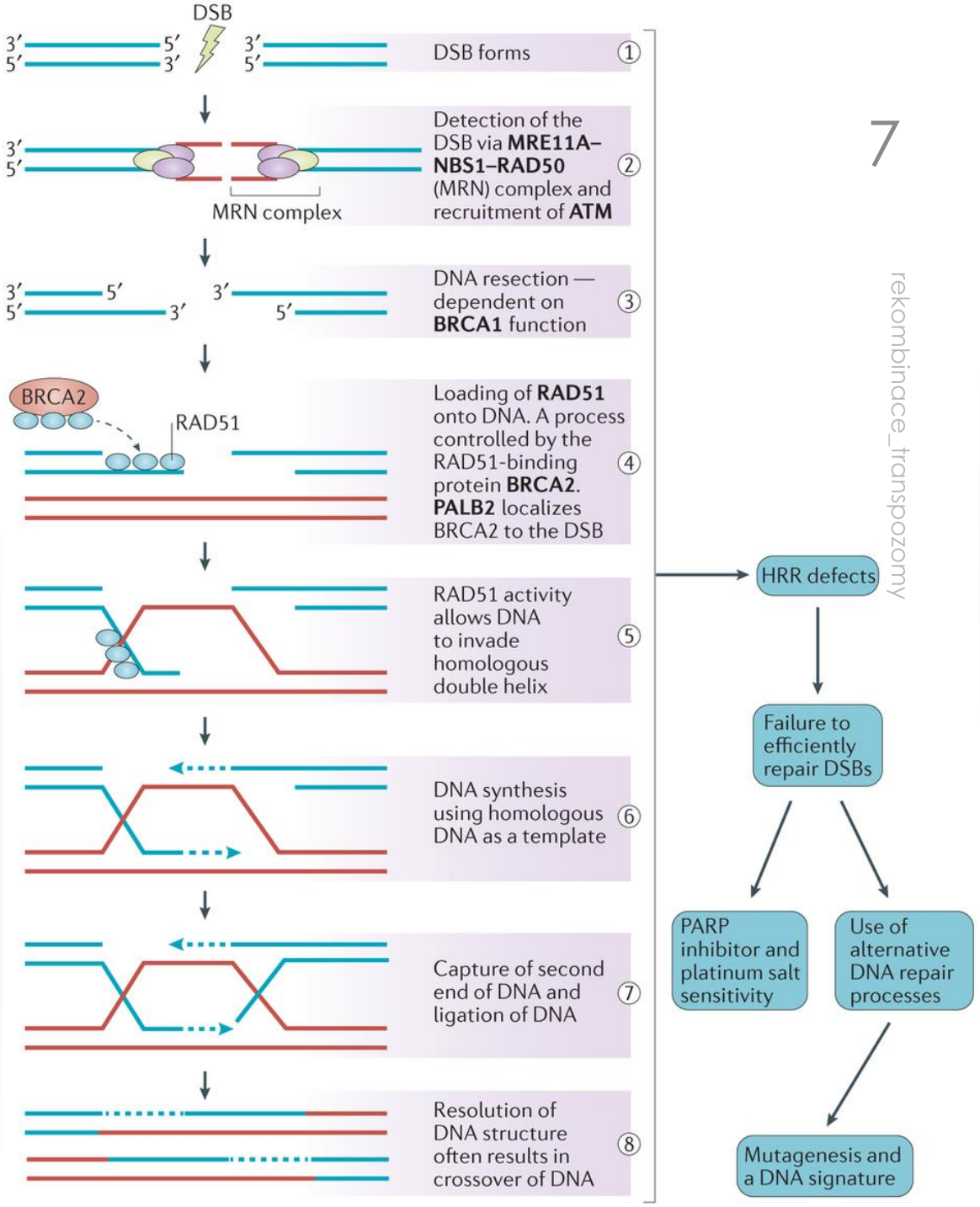
separation of recombinant molecules by **resolvase**

Recombination during meiosis



Repair of double-strand breaks

MRE11
RAD50
BRCA1
BRCA2



Site-specific recombination

- recombination between non-homologous sequences
- mechanism whereby the genome moves mobile genetic elements
- controlled by enzymes which recognize short sequences at the ends of the mobile elements, does not require extensive DNA homology

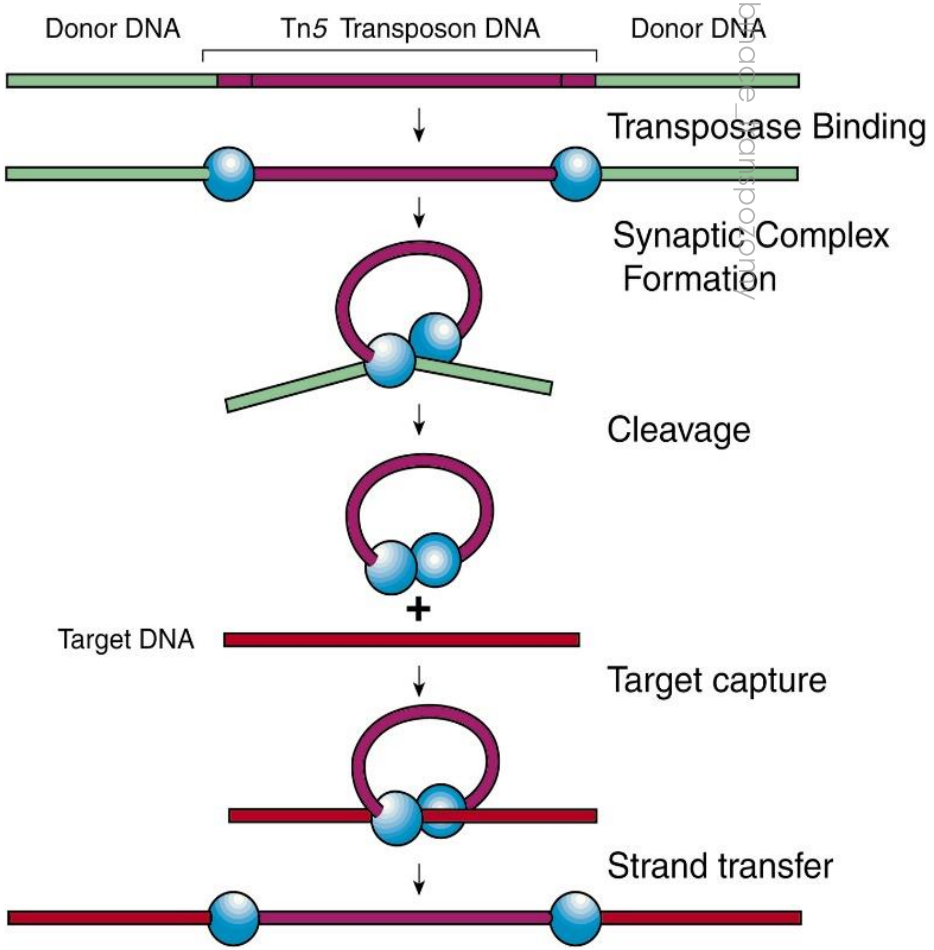
Transposons, mechanisms of transposition, retroelements

Transposons = mobile genetic elements

- Cause **changes in genetic information** (insertional inactivation, activation, modulation of gene expression, mutation)
- **Significantly involved in the architecture of genome** - most of repetitive DNA in the genome of the plant consists of transposon sequences
- In plants, transposons **do not carry genes** which directly increase the fitness (unlike many bacterial transposons)

Transposons = mobile genetic elements

- segments of DNA capable of transfer to another site of the genome **(transposition)**
- in all prokaryotes and eukaryotes yet analyzed (with the exception of the parasite *Plasmodium falciparum*)
- **Transposase** enzyme
- They do not exist as a separate independent form like plasmids or phages
- **important source of genomic instablity**
- - In plants - thousands of families (**80% of the genome)**
- - animals 3-45%,
- - fungi 2-20%
- - human **40%**



The discovery of transposons

Demerec (1937) described the unstable mutations in the D. melanogaster
B. McClintock (Nobel Prize 1983) showed in the 40s and 50s while studying chromosomal breaks in maize that its genome contains many mobile elements causing somatic mutations (ac / ds)
Molecular analysis of these elements may be implemented up roughly from the late 70s ,the first cloned elements are elements of D. melanogaster (1978), which are now known as "Copia-like" elements
Transposons are found in all organisms, which have been searched, (except parasite Plasmodium falciparum) in plants up to 80% of the genome, in animals 3-45%, fungi 2-20%
Transposons are segments of DNA capable of transfer to another place of the genome (transposition), either autonomously or with the help of related elements.

- **Barbara McClintock (1902-1992)**
- The Nobel Prize in Physiology and Medicine in 1983 for discovering (knowledge of the nature) of mobile genetic elements in maize
- Study of chromosomal breakage in maize
- increased incidence of breaks in a certain area (= a marker called "dissociation" Ds)
- position of marker was not stable after crossing with some lines, and shifted to other spots (= line carrying the "activator" Ac)

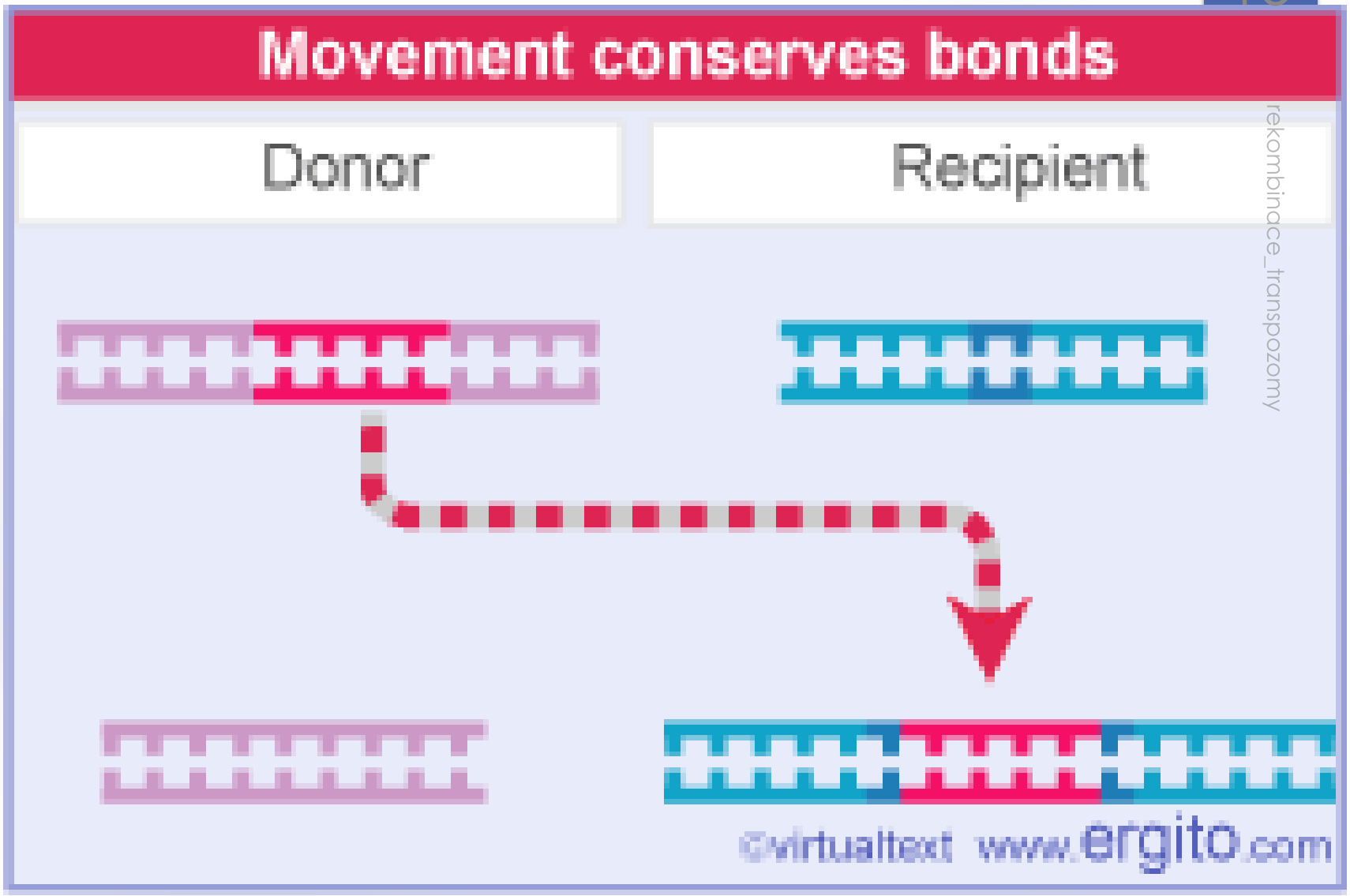
- Barbara McClintock discovered the first TEs in maize (*Zea mays*) at the Cold Spring Harbor Laboratory in New York. McClintock was experimenting with maize plants that had broken chromosomes.[5]
- In the winter of 1944–1945, McClintock planted corn kernels that were self-pollinated, meaning that the silk (style) of the flower received pollen from its own anther.[5] These kernels came from a long line of plants that had been self-pollinated, causing broken arms on the end of their ninth chromosomes.[5] As the maize plants began to grow, McClintock noted unusual color patterns on the leaves.[5] For example, one leaf had two albino patches of almost identical size, located side by side on the leaf.[5] McClintock hypothesized that during cell division certain cells lost genetic material, while others gained what they had lost.[6] However, when comparing the chromosomes of the current generation of plants with the parent generation, she found certain parts of the chromosome had switched position.[6] This refuted the popular genetic theory of the time that genes were fixed in their position on a chromosome. McClintock found that genes could not only move, but they could also be turned on or off due to certain environmental conditions or during different stages of cell development.[6]
- McClintock also showed that gene mutations could be reversed.[7] She presented her report on her findings in 1951, and published an article on her discoveries in *Genetics* in November 1953 entitled "Induction of Instability at Selected Loci in Maize."[8]
- Her work would be largely dismissed and ignored until the late 1960s-1970s when it would be rediscovered after TEs were found in bacteria.[9] She was awarded a Nobel Prize in Physiology or Medicine in 1983 for her discovery of TEs, more than thirty years after her initial research.[10]
- Approximately 90% of the maize genome is made up of TEs,[11] as is 44% of the human genome.[12]

Types of transposition

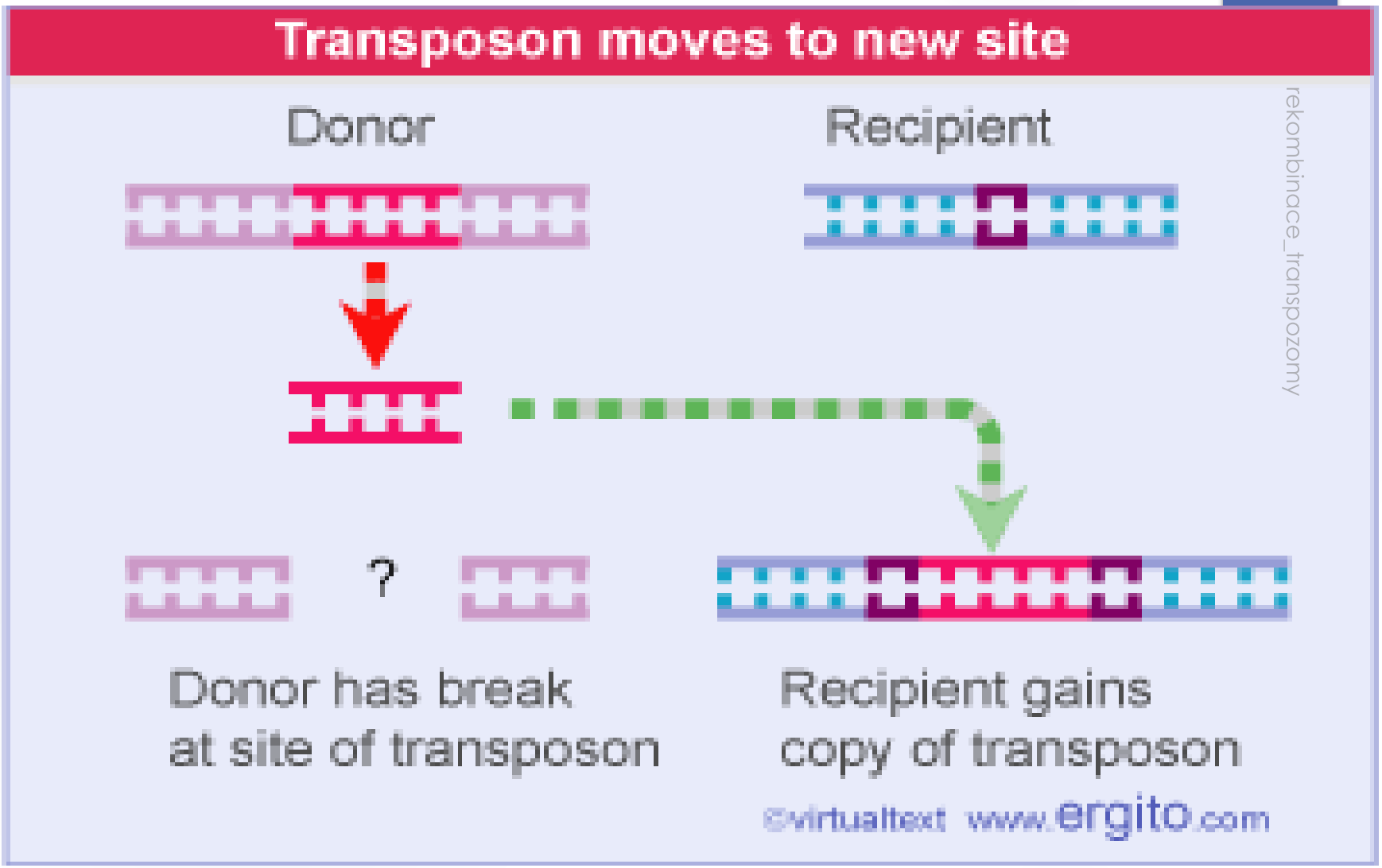
- **Class I -conservative transposition - „CUT and PASTE“**
  - transposon excision and transfer to another place of the genome
  - Only transfer, without multiplication

- **Class II-replicative transposition - „COPY and PASTE“**
- replication, the copy is placed in a new location
- original element remains, the number of copies ~ number of replications
- copying through RNA intermediate or direct insertion of copied DNA

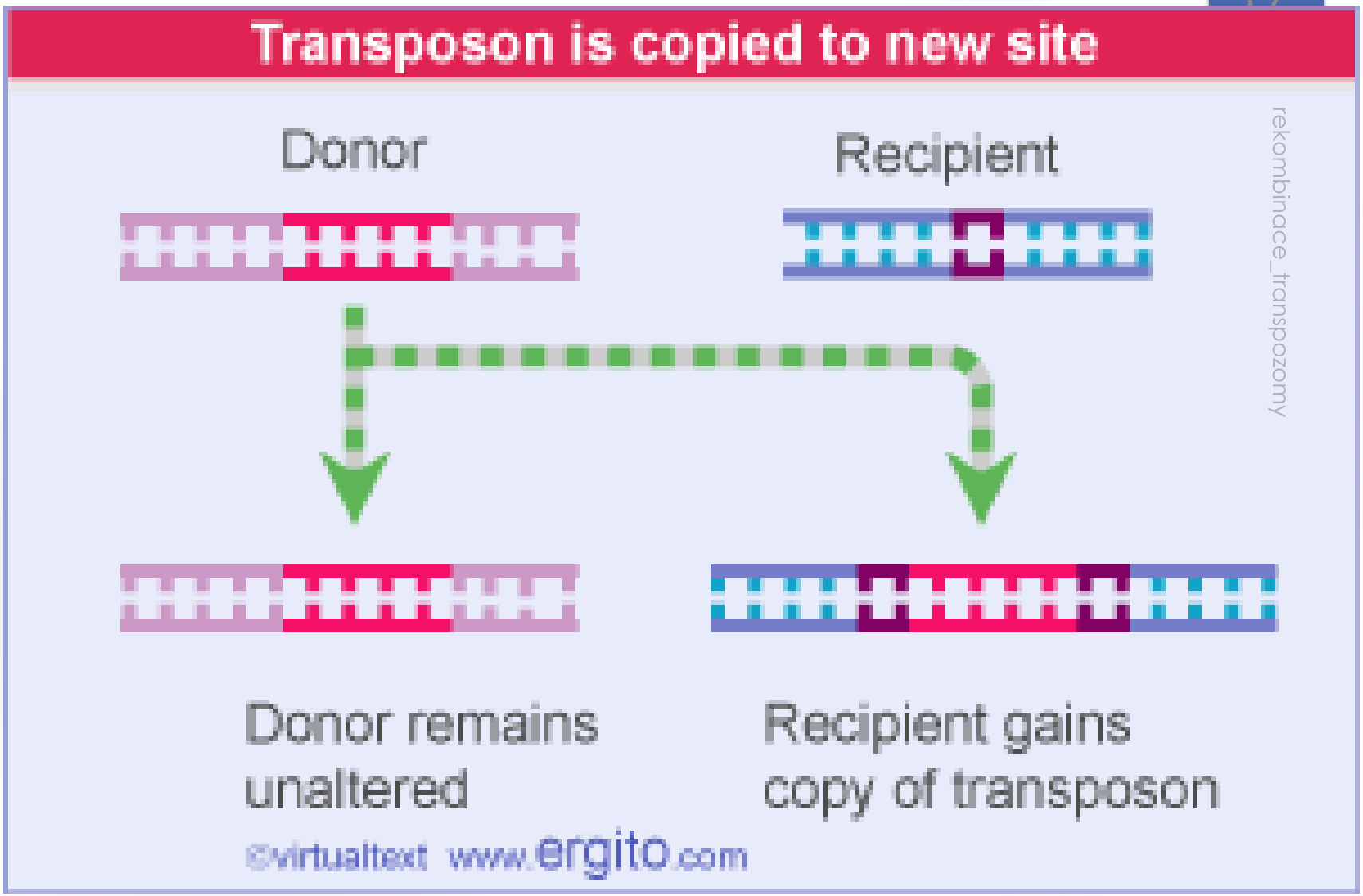
Conservative x duplicative transposition



Conservative x duplicative transposition



Conservative x duplicative transposition



Classification of transposons

1. **Class:** **Depending on whether or not with RNA intermediate**
   - **DNA transposons**
   - **Retrotransposons**

2. **Subclass:** **According to the mechanism of replication (for DNA transposons)**
3. **Order:** **According to the basic structural features**
4. **Superfamily:** **By sequence homology**

A unified classification system for eukaryotic transposable elements

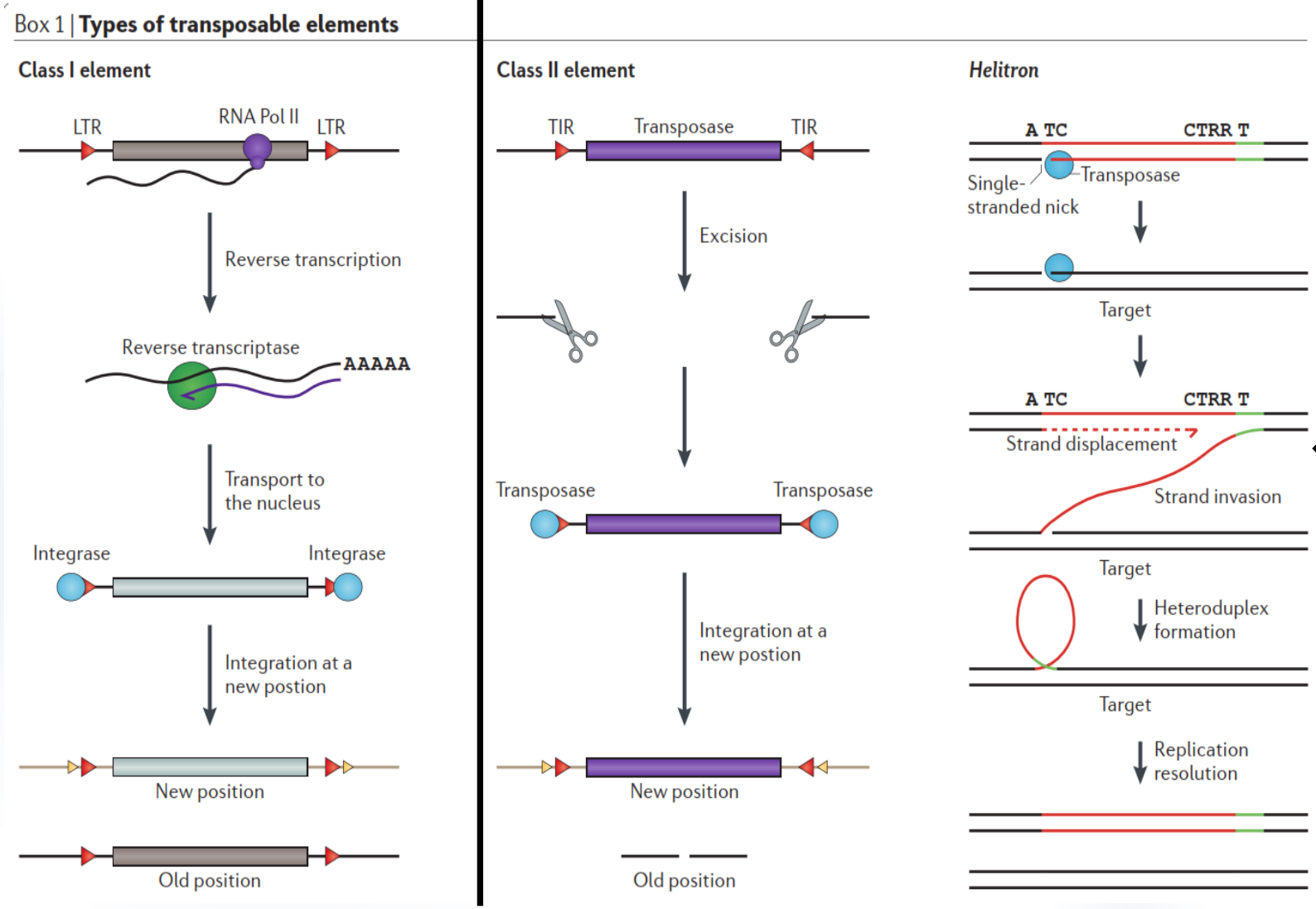
Thomas Wicker, François Sabot, Aurélie Hua-Van, Jeffrey L. Bennetzen, Pierre Capy, Boulos Chalhoub, Andrew Flavell, Philippe Leroy, Michele Morgante, Olivier Panaud, Etienne Paux, Phillip SanMiguel and Alan Schulman

Nature Rev. Genet. 2008

Basic types of TE

Retrotransposons

DNA transposons



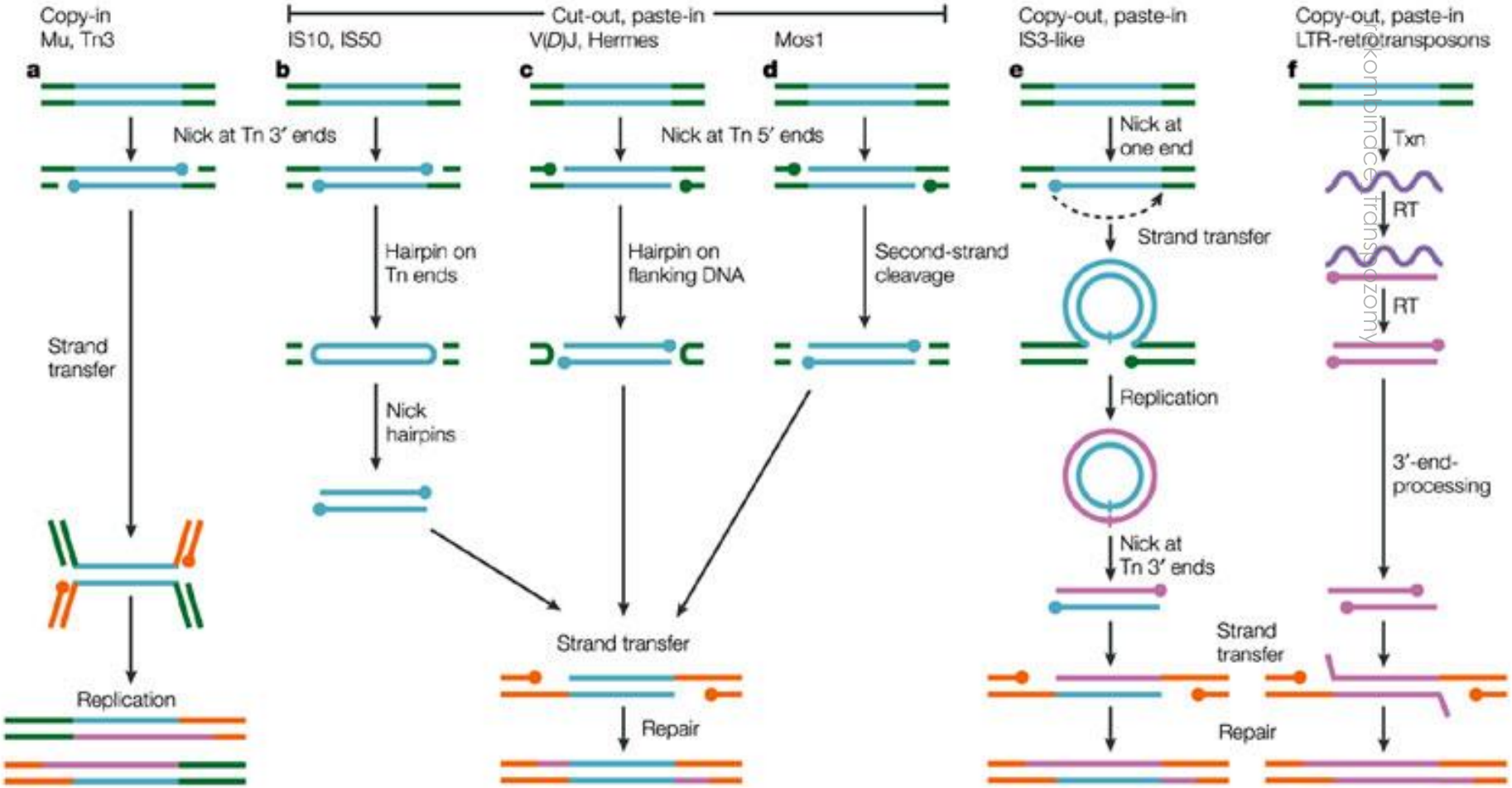
rekombinace_transpozom

Rolling circle

replication ???

Lisch 2013, Nature Rev. Genet.

Types of transposition



Types of transposable elements (Not all transposons encode the necessary enzymatic activity)

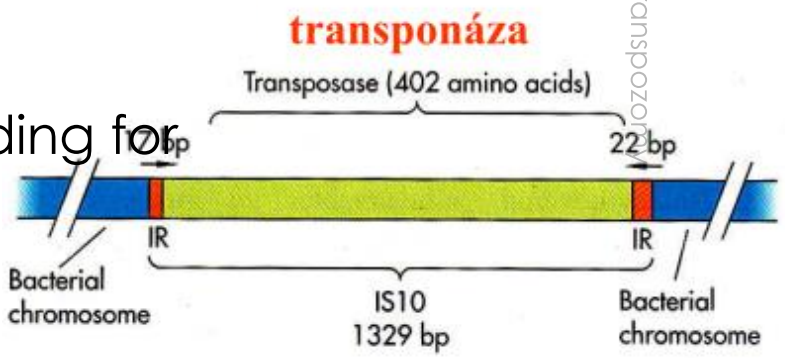
- **autonomous elements**
  - – encoding the **gene** whose product ensures transposition / replication

- **non-autonomous elements**

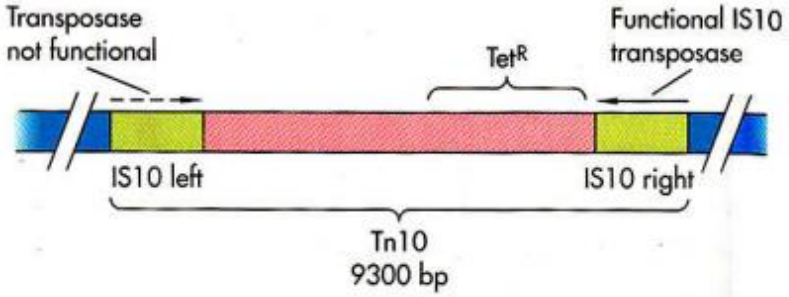
  – derived from autonomous
    - – lost the genes required for the transposition, but can be mobilized by other related autonomous elements
    - – have cis sequences necessary to mobilize

A) Transposable elements in bacteria

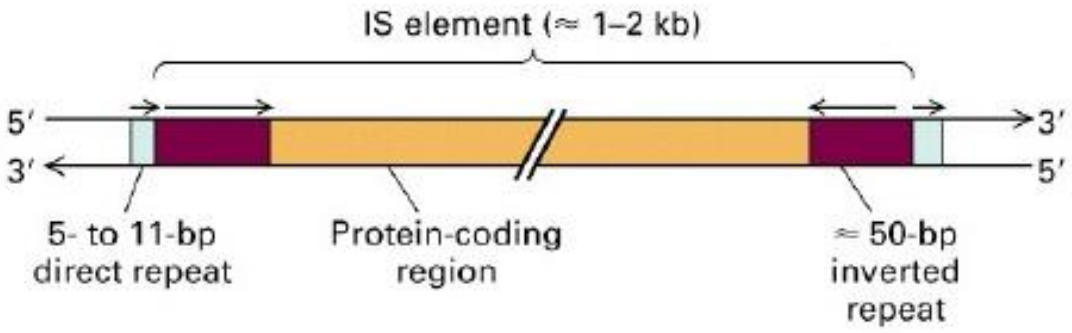
- First TE studied at the molecular level
- relocated within the bacterial chromosome or the chromosome and plasmid
- The main types:
- **IS-elements** (only include genes coding for proteins providing transposition)



- **composite transposons and Tn3-elements** (additionally contain genes which encode products functionally unrelated to the process of transposition)

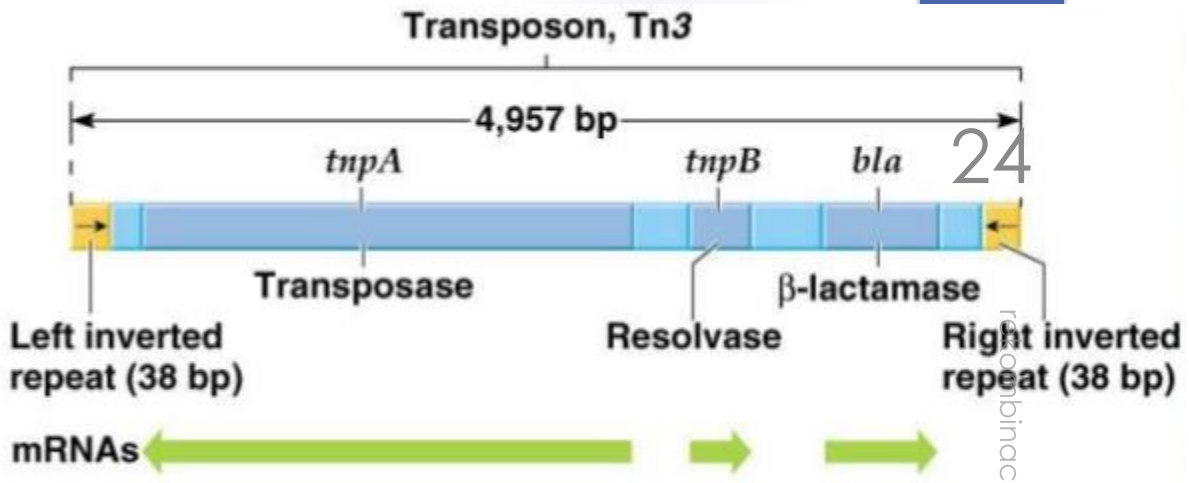


IS-elements



- usually less than 2500 bp
- framed by short identical sequences - **inverted terminal repeats**
- mutations in terminal repeats eliminates transposition capability
- They contain only genes for ensuring and controlling transposition
- encode **transposase** enzyme: binds to the ends of the element, cleaves both DNA strands - thus the element is released from the original site

Transposon Tn3



- It contains genes that are not necessary for transposition
- the ends are formed by simple inverted repeats
- at the target site duplication occurs

- Structure:
- **transposase/resolvase gene and their repressor**
- **gene for beta-lactamase (Amp resistance)**

Importance of bacterial transposons in medicine

- often they contain genes for resistance to antibiotics
- that these genes can spread easily and thereby increase the resistance of pathogenic bacteria to antibiotics
- today it is difficult to treat a variety of infectious diseases (diabetes, gonorrhea, tuberculosis, etc.).
- spread of resistance is promoted by the widespread use of antibiotics
- transposons (transfer between the molecules of DNA within the bacterial cells) and conjugative plasmids (transfer between different bacterial strains)

Kanamycin

Gentamycin

Ampicilin

Tetracyklin

Chloramfenikol

Streptomycin

Bacterial transposons: summary

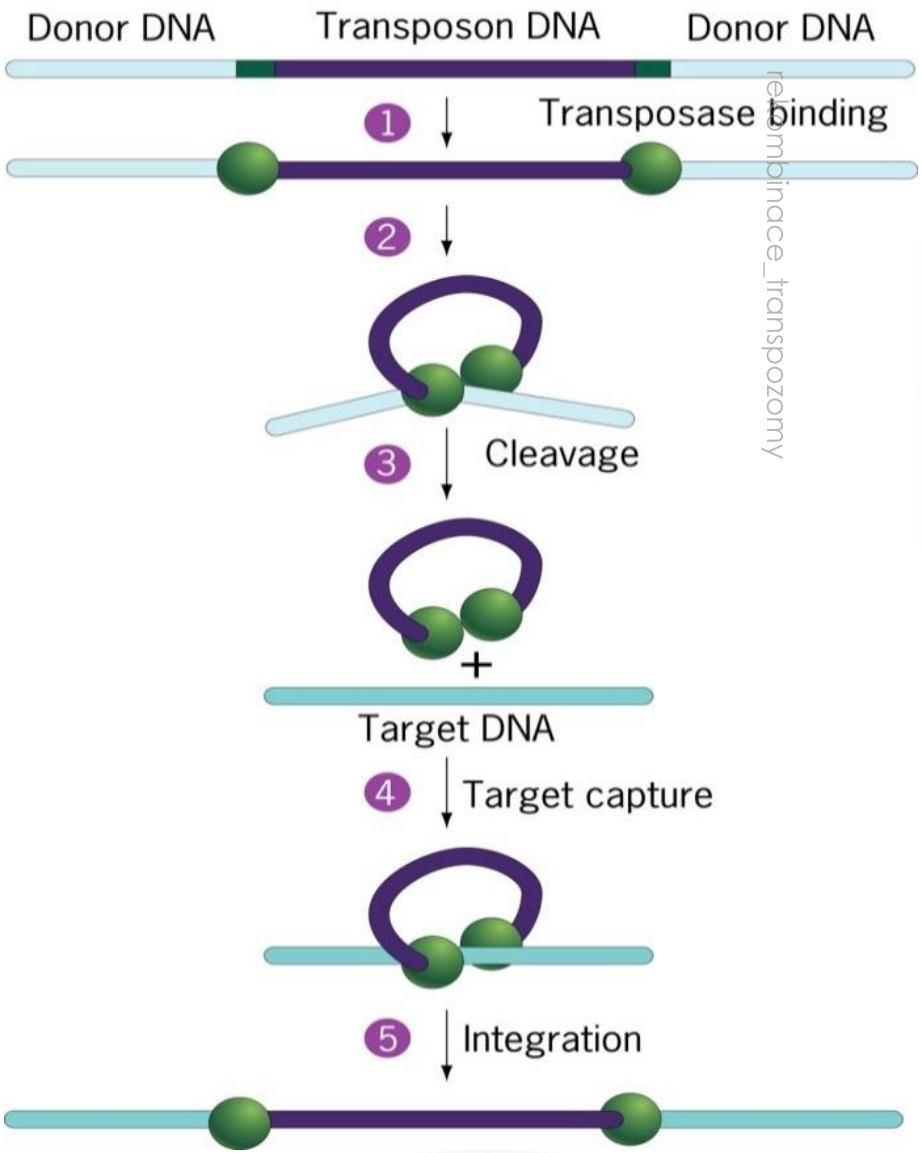
- Insertion sequences - IS-elements, "cut and paste" transposon, part of bacterial chromosomes and plasmids
- Composite transposons generated by 2x IS-elements, flank area for one or more genes for resistance to antibiotics
- Tn3-type replicative transposon, temporarily connects the molecules to form co-integrate, when unfolded, each molecule contains 1xTn3
- Bacterial transposons bounded by inverted repeats are duplicated after incorporation
- Conjugative plasmids - carrying a transposon containing the resistance genes from one bacterium to another

B) Transposons in eukaryotes

- **mainly types of "cut and paste" and retrotransposons**
- **P-elements** in Drosophila
- **Ty-elements** in yeast
- **human retrotransposons LINE** constitute about 15% of the genome (mostly immobile due to mutations - incapable of transposition)
- some can maintain mobility and can cause **diseases** (e.g. transposition into the gene for a factor required for blood clotting - hemophilia)

DNA transposons - subclass I:

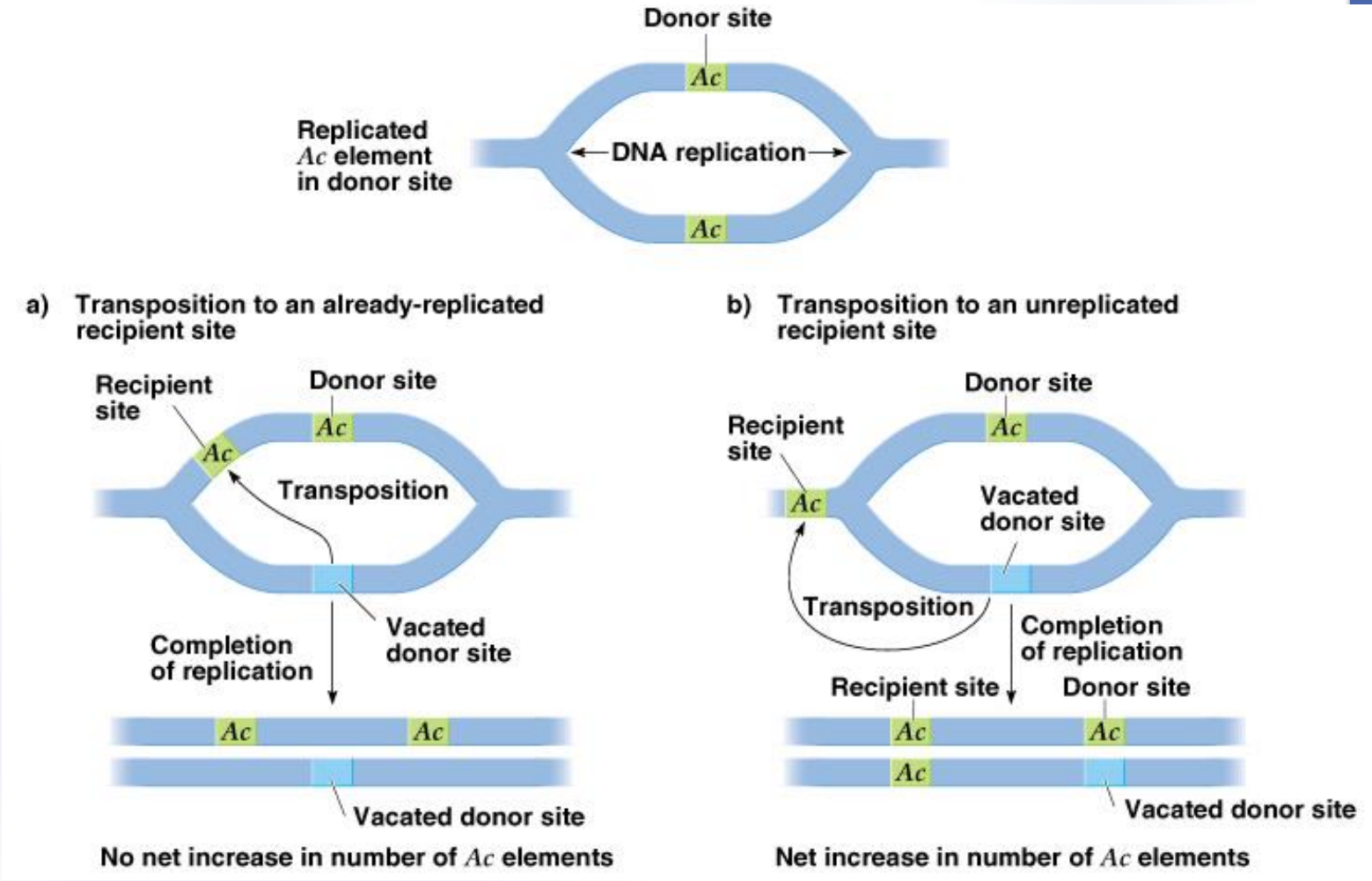
- encode **transposase**, the edges are inverted repeats
- transposition - complex process – binding of IR, cleavage (transposase), cleavage of the target sequence, DNA synthesis, ligation
- duplication of short sequence (2-8 bp) in incorporation site = footprint after re-excision



DNA transposons - subclass I:

- usually integration in the vicinity of the original insertion
- usually a few to a few hundred copies in the genome
- Ac, Spm, Mu (maize), Tam (Antirrhinum), Tphl (petunia) Tags (Arabidopsis), Stowaway, Tourist> 10,000 copies every 30 kbp (maize, insertion into the TA-rich sequences)
- MULE (Mutator-like elements) with rice - over 1,000 gene fragments mobilized - 5% is expressed - evolution of new genes
- mutated **non-autonomous** forms Ac/**Ds (Ds1**, **Ds2)**, Spm/**dSpm**

Movements and propagation of DNA transposons

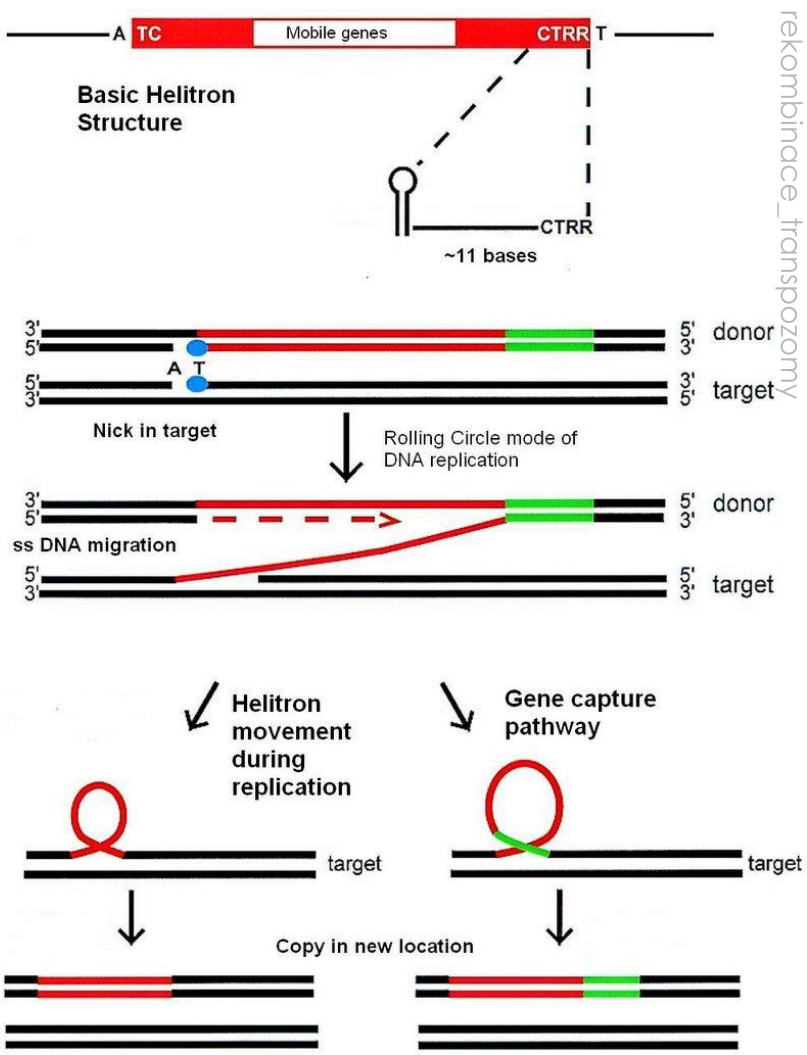


The activation mechanism during replication? **Hemimethylated status?**

break repair after TE excision by homologous second chromatid section (possibility of reconstructing the original sequence with TE = amplification)

DNA transposons - subclass II:

- order: Helitron - a single-stranded break, DNA migration, insertion
- in maize 4-10 thousands of mobile genetic elements



RETROTRANSPOZONS

Class I TEs are copied in two stages: first, they are transcribed from DNA to RNA, and the RNA produced is then reverse transcribed to DNA. This copied DNA is then inserted back into the genome at a new position.

The reverse transcription step is catalyzed by a reverse transcriptase, which is often encoded by the TE itself. The characteristics of retrotransposons are similar to retroviruses, such as HIV.

Retrotransposons are commonly grouped into three main orders:

- TEs with **long terminal repeats (LTRs),** which encode reverse transcriptase, similar to retroviruses
- **Long interspersed nuclear elements (LINEs, LINE-1s, or L1s),** which encode reverse transcriptase but lack LTRs, and are transcribed by RNA polymerase II
- **Short interspersed nuclear elements (SINE)** do not encode reverse transcriptase and are transcribed by RNA polymerase III

[Note : Retroviruses can also be considered TEs. For example, after conversion of retroviral RNA into DNA inside a host cell, the newly produced retroviral DNA is integrated into the genome of the host cell. These integrated DNAs are termed *proviruses*. The provirus is a specialized form of eukaryotic retrotransposon, which can produce RNA intermediates that may leave the host cell and infect other cells. The transposition cycle of retroviruses has similarities to that of prokaryotic TEs, suggesting a distant relationship between the two].

Retrotransposons

Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
Class I (retrotransposons)					
LTR	*Copia*	GAG AP INT RT RH	4–6	RLC	P, M, F, O
	Gypsy	GAG AP RT RH INT	4–6	RLG	P, M, F, O
	Bel–Pao	GAG AP RT RH INT	4–6	RLB	M
	Retrovirus	GAG AP RT RH INT ENV	4–6	RLR	M
	ERV	GAG AP RT RH INT ENV	4–6	RLE	M
DIRS	*DIRS*	GAG AP RT RH YR	0	RYD	P, M, F, O
	Ngaro	GAG AP RT RH YR	0	RYN	M, F
	VIPER	GAG AP RT RH YR	0	RYV	O
PLE	*Penelope*	RT EN	Variable	RPP	P, M, F, O
LINE	*R2*	RT EN	Variable	RIR	M
	RTE	APE RT	Variable	RIT	M
	Jockey	ORF1 APE RT	Variable	RIJ	M
	L1	ORF1 APE RT	Variable	RIL	P, M, F, O
	I	ORF1 APE RT RH	Variable	RII	P, M, F
SINE	*tRNA*		Variable	RST	P, M, F
	7SL		Variable	RSL	P, M, F
	5S		Variable	RSS	M, O

Structural features

Long terminal repeats; Terminal inverted repeats; Coding region; Non-coding region; Diagnostic feature in non-coding region; Region that can contain one or more additional ORFs

Protein coding domains

AP, Aspartic proteinase; APE, Apurinic endonuclease; ATP, Packaging ATPase; C-INT, C-integrase; CYP, Cysteine protease; EN, Endonuclease; ENV, Envelope protein; GAG, Capsid protein; HEL, Helicase; INT, Integrase; ORF, Open reading frame of unknown function; POL B, DNA polymerase B; RH, RNase H; RPA, Replication protein A (found only in plants); RT, Reverse transcriptase; Tase, Transposase (* with DDE motif); YR, Tyrosine recombinase; Y2, YR with YY motif

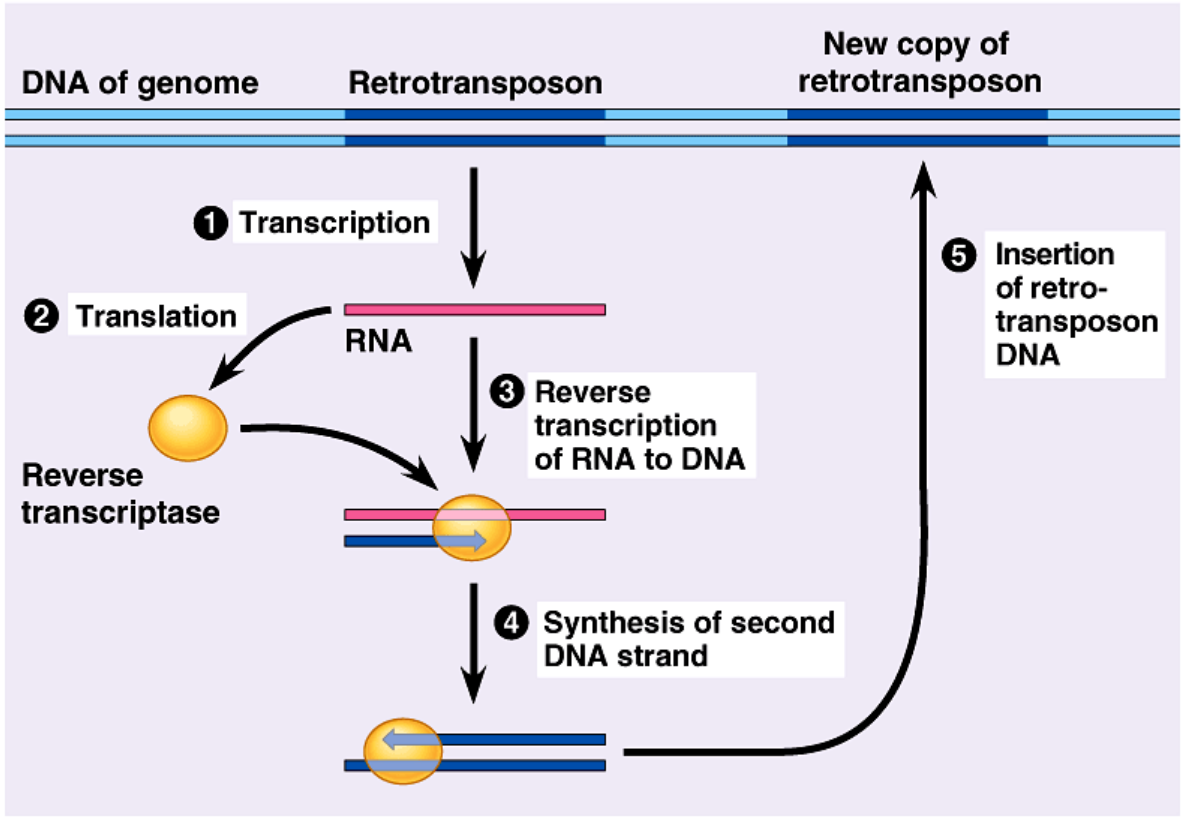
Species groups

P, Plants; M, Metazoans; F, Fungi; O, Others

Nature Reviews | Genetics

Retrotransposons

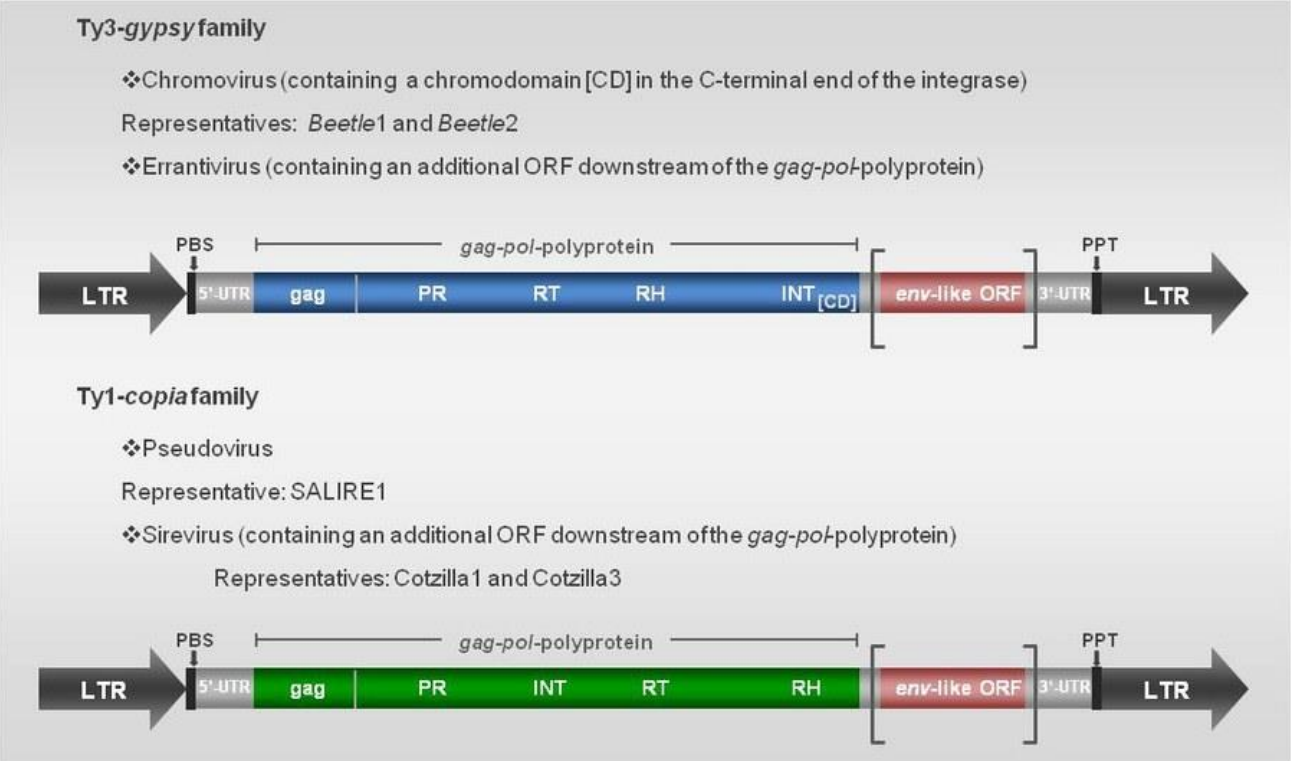
- **replication** through RNA intermediate
- 1-13 kbp size (apart form SINE), millions of copies (up to 40-80% of the genome)
- often in heterochromatic regions, in euchromatin especially among genes - possibly as a result of selection pressure



Retrotransposons – elements similar to retroviruses

Retrotranspozons

- Odrer: LTR
  - LTR (long terminal repeat): promoter, terminator, direct repeat
  - short duplication of the target sequence
  - - **protease, reverse transcriptase, RNase H, integrase, nucleocapsid protein**



LTR retrotransposons

- **Ty1- copia group**

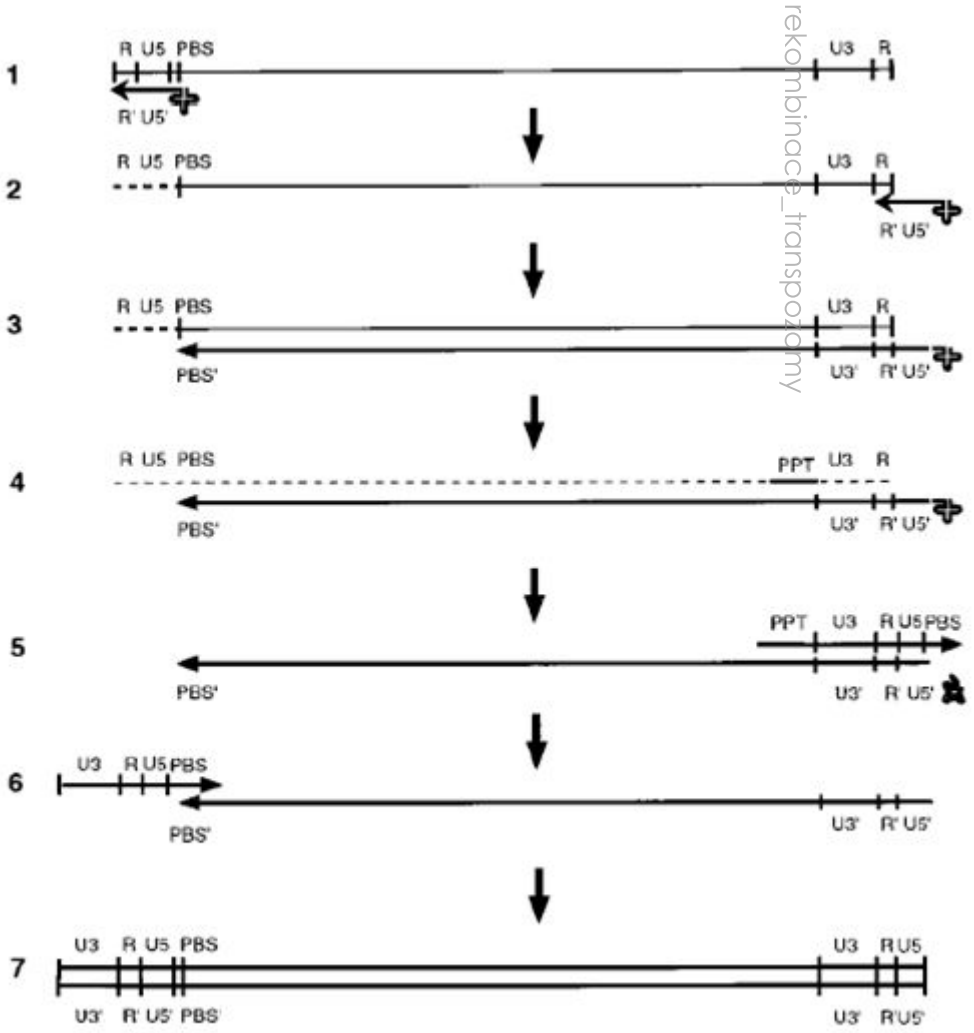
- BARE-1 barley, 12.1 kbp> 50,000 copies of transcript in leaves and callus
- Opie -1, maize, 8.7 kbp,> 30,000 copies, roots, leaves, integration into the LTR
- PREM-2, maize, 9.5 kbp,> 10,000 copies in microspore
- TNT1, tobacco, 5.3 kbp,> 100, protoplasts, roots, - activation after injury, pathogen attack, integration into euchromatin

- **Ty3 – gypsy group**
- potential ancestors of animal retroviruses, sometimes env-like sequences

- Athila, A.t., 10,5 kbp, >10000, paracentromeric regions
- Athila-1-1, A.t., 12 kbp, 730, env-like sequences
- Cinful-1, maize, 8,6 kbp, 20000, leavse, env-like seq.

LTR retrotransposons - replication

- - replication analogous to retroviruses - LTR (U3, R, U5)
- - PBS (primer binding site): tRNA primer
- - jumps between templates
- (direct repeat - R)



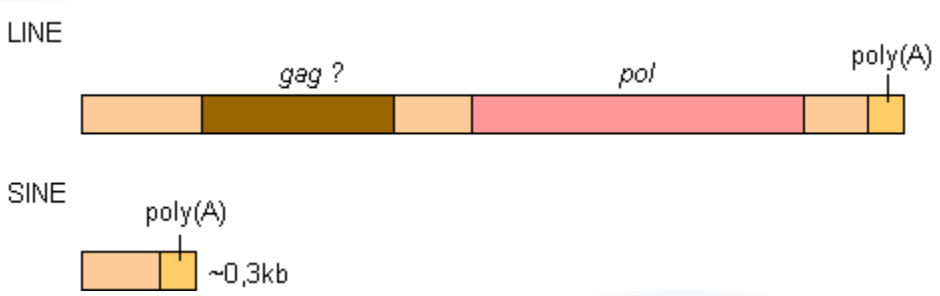
Transposable elements in humans

- 44% of human DNA is derived from transposable elements
- many different types:
- **long dispersed nuclear elements (LINE)** - size about 6 kb
- **short dispersed nuclear elements (SINE)** - less than 400 pb
- use of reverse transcription

Retrotransposons w/o LTR

- **LINE (long interspersed nuclear elements)**
- - apparently phylogenetically oldest predecessor of transposons with LTR
- - 5´end – promoter; 3´end - terminator
- Cin4, maize, 1-6,8kbp, 50-100, variously truncated forms

- **SINE (short interspersed nuclear elements)**
- - using RT apparatus of other transposons (non-autonomous)
- derived from RNA polymerase III products (tRNA 7SLRNA, rRNA)
- < 500 nt



Regulation of the activity of transposons

- **Retrotransposons**
- enormous potential to change gene function and genome structure
- regulation by own control mechanisms and host (mostly inactive - methylation, controlled activation developmentally, external conditions)
- coevolution of mechanisms regulating transposition, insertion specificity, mutagenic potential
- functions: changes in gene regulation, role in DNA repair, centromeres

- **DNA transposons**
- regulation of activity by environmental conditions:
- Tam1 in snapdragon (1000 * at 15 ° C)
- Methylation

Methylation of transposons

- Inactivation - temporary, permanent, the possible cause of methylation mechanisms
- For retrotransposons - similarities with the silencing of multiple-copy genes
- The activity of *Ac* and *Spm* is different depending on the type of gametes (changes during gametogenesis)
- The increase of *Spm* and *Mu* methylation with development (leaves), demethylation in the early stages of development
- Methylation is needed especially during meiosis – safeguarding of integrity (x illegitimate crossover)

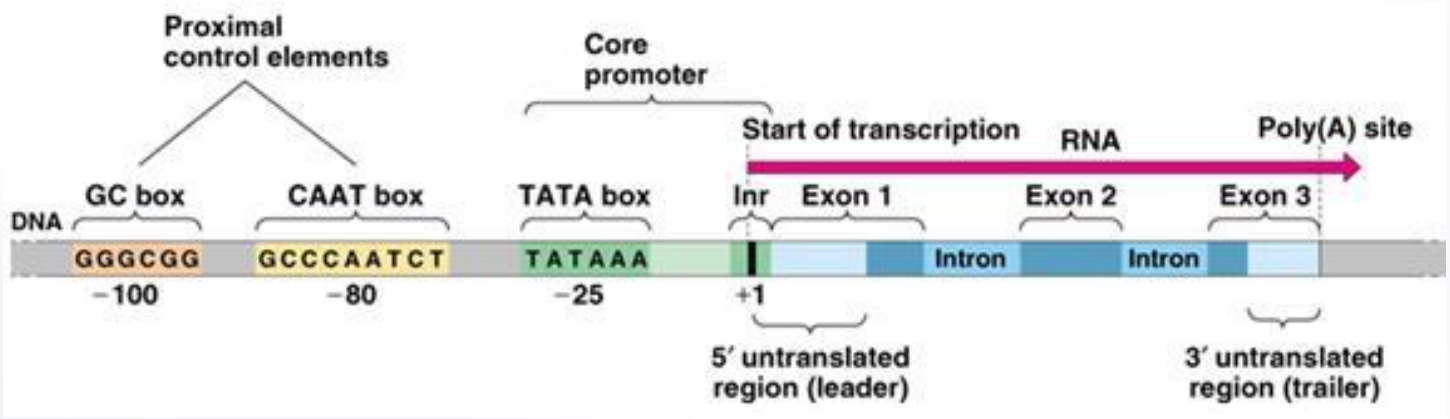
Significance of transposons

- **inducing mutations = increasing variability**
- **modulating the expression (activation, repression, during development, during stress)**
- **creation of new genes**
- **induction of chromosomal rearrangements induced by recombination between transposons**

- in plants transposons do not carry genes which directly increase the fitness (resistance etc.).
- increasing fitness by randomly induced mutations (e.g. activation by stress conditions) - very low probability ...
- great importance in the domestication (breeding) of plants

Mutations caused by transposons

- place of incorporation (different preferences: GC, AT)
- character of carried regulatory sequences



- Modulation of expression (time and place) - **promoter, enhancer**
- changes in the stability of the transcript and posttranscriptional editing (splicing) - **UTR, introns, terminator**
- change in the sequence of the resulting protein, premature termination of translation, creation of chimeric genes, ... - **exons, introns**

Regulation of gene expression by transposons

- prevention or reduction of transcription
- modulation of time and site specific expression
- changes in the stability of the transcript and posttranscriptional editing (splicing)
- change in the structure of the resulting protein

- e.g. Maize- inactivation of the gene CCT (response to photoperiod length) by inserting a cacti-like element (TE DNA) in the promoter region - expansion of cultivation in temperate zones (flowering during a long day)

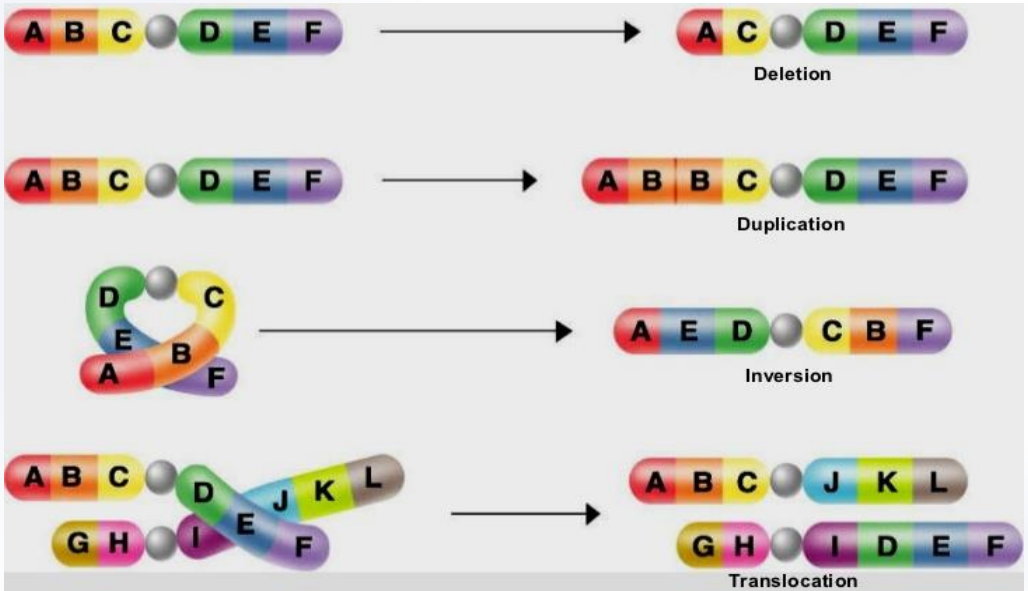
Significance in the evolution of genes

- insertional mutagenesis (premature termination)
- possible participation in the multiplication of genes
  - directly or indirectly via homologous recombination
  - advantageous to have a gene family of different regulations, respectively backup copies of genes
- creation of intron-free copies of genes (reverse transcription)
- may participate in the creation of entirely new genes - eg. fusion of transmitted fragments of existing genes (helitrons, MULE)
- genes that were originally of transposon origin were "domesticated" by many eukaryotic organisms for new features (eg. telomerase, syncitin, ...)

- **natural genetic engineering tools**
- **they spread and thus can provide a selective advantage for the host**
- **others are genetic parasites**

Changes on the level of the genome

- possible participation in the multiplication of genes
- creation of intron-free copies of genes
- chromosomal rearrangements (repetitive sequences)
- breaks, inversions, deletions, duplications, translocations, ...



TE in disease

- TEs are mutagens and their movements are often the causes of genetic disease. They can damage the genome of their host cell in different ways:[27]
- a transposon or a retrotransposon that inserts itself into a functional gene will most likely disable that gene;
- after a DNA transposon leaves a gene, the resulting gap will probably not be repaired correctly;
- multiple copies of the same sequence, such as Alu sequences, can hinder precise chromosomal pairing during mitosis and meiosis, resulting in unequal crossovers, one of the main reasons for chromosome duplication.
- **Diseases often caused by TEs include hemophilia A and B, severe combined immunodeficiency, porphyria, predisposition to cancer, and Duchenne muscular dystrophy.[28][29] *LINE1* (*L1*) TEs that land on the human Factor VIII have been shown to cause haemophilia[30] and insertion of *L1* into the *APC* gene causes colon cancer, confirming that TEs play an important role in disease development.[31]**
- Additionally, many TEs contain promoters which drive transcription of their own transposase. These promoters can cause aberrant expression of linked genes, causing disease or mutant phenotypes.

Discovery of transposons

- **Barbara McClintock (1902-1992)**
- The Nobel Prize in Physiology and Medicine in 1983 for discovering (knowledge of the nature) the mobile genetic elements in maize
- Study of chromosomal breakage in maize
- increased incidence of breaks in a certain area (= a marker called "dissociation" Ds)
- position of marker was not stable after crossing with some lines, and shifted to other spots (= line carrying the "activator" Ac)

Discovery of transposons

- in one line a Ds marker shift caused a loss of purple discoloration of the caryopses
- light color caryopses (c) caused by the insertion of the Ds element were not stable in a crossing with lines carrying Ac - appearance of caryopses with purple spots
- triploid endosperm

- c/c/c = light color
- C/c/c or C/C/c pr C/C/C = purple color

Transposition and coloration of the caryopses

- If the c is reversed to C, red pigment begins to form in the cell thus forming a spot on a light background
- the earlier in the development of the caryopsis reversion occurs, the greater the stain
- B. McClintock concluded that „c" allele was created by integrating the non-autonomous transposon "Ds" to "C" allele (Ds = dissociation)
- reversion of c to C is due to the transposition of the Ds element from the C allele which is mediated by autonomous transposable element
- "Ac" (Ac = activator)