MUNI
SCI

Bi4025en
Molecular Biology

Mgr. Jiří Kohoutek, Ph.D.

Lecture 11

- Mobile genetic elements, transposons and retrotransposons

Maize – *Zea Mays*

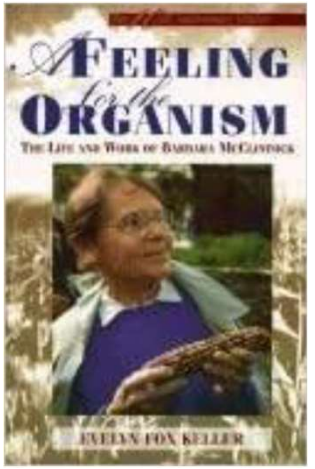
- One of the world's most important crops.
- The natives of the Americans cultivated many different and differently colored varieties, attributing aesthetic and religious significance to colors.
- Scientific significance: color patterns are the result of a phenomenon called transposition.

- Barbara McClintock was awarded by Nobel Prize in Physiology or Medicine 1983.
- Careful observation of plants in the field.
- Changes in inheritance ratios and traits such as grain color.

MUNI SCI

Barbara McClintock

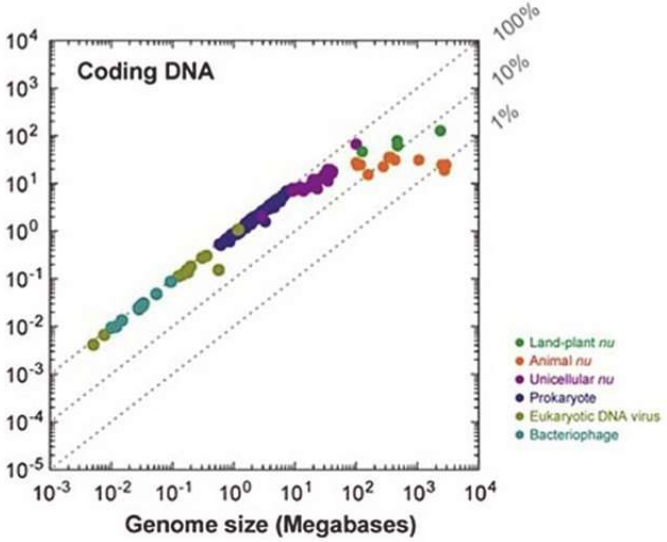
- Transposable elements were discovered by Barbara McClintock during experiments conducted in 1944 on maize.
- Since they appeared to influence phenotypic traits, she named them controlling elements.
- However, the idea that a gene could "jump" from place to place was contrary to Mendel's laws.
- Her presentation at the 1951 Cold Spring Harbor Symposium was not understood and at least not very well received. She had no better luck with her follow-up publications.
- Her discovery was brought back to life after discovery of insertion sequences (IS) in bacteria by Szybalski´s group in the early1970s.

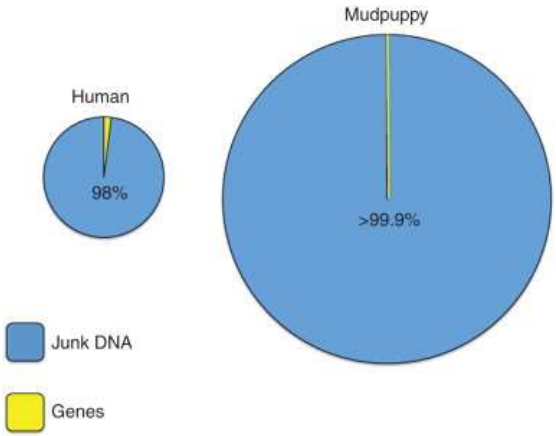


Methods in Molecular Biology, vol. 1910, https://doi.org/10.1007/978-1-4939-9074-0_6

C – value paradox

- Rejecting the idea of an static genome meant changing the paradigm of genetics.
- Part of the new view was the acceptance that genomes are largely made up of repeating sections of DNA (repetitive DNA).





- C-value paradox in Eukaryotes: increase of genome size depends of the increase of non-coding elements.
- The human genome, for example, comprises less than 2% protein-coding regions.

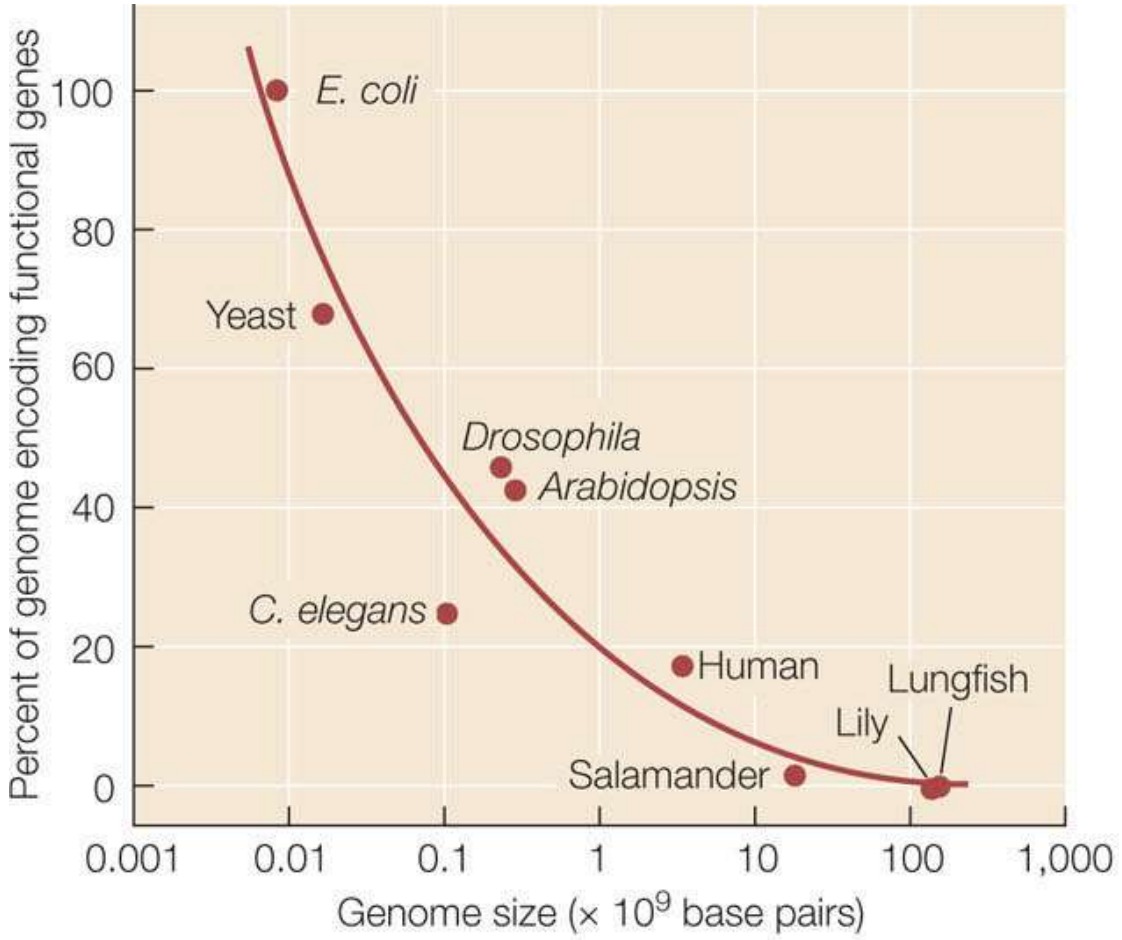
https://www.sciencedirect.com/topics/neuroscience/c-value

The Role of Non-coding Structural Information in Phylogeny, Evolution and Disease, DOI: 10.13140/RG.2.1.3803.6322

MUNI SCI

C – value paradox

- A Large proportion of DNA is noncoding.
- Most of the DNA of bacteria and yeasts encodes RNAs or proteins, but a large percentage of the DNA of multicellular species is noncoding.

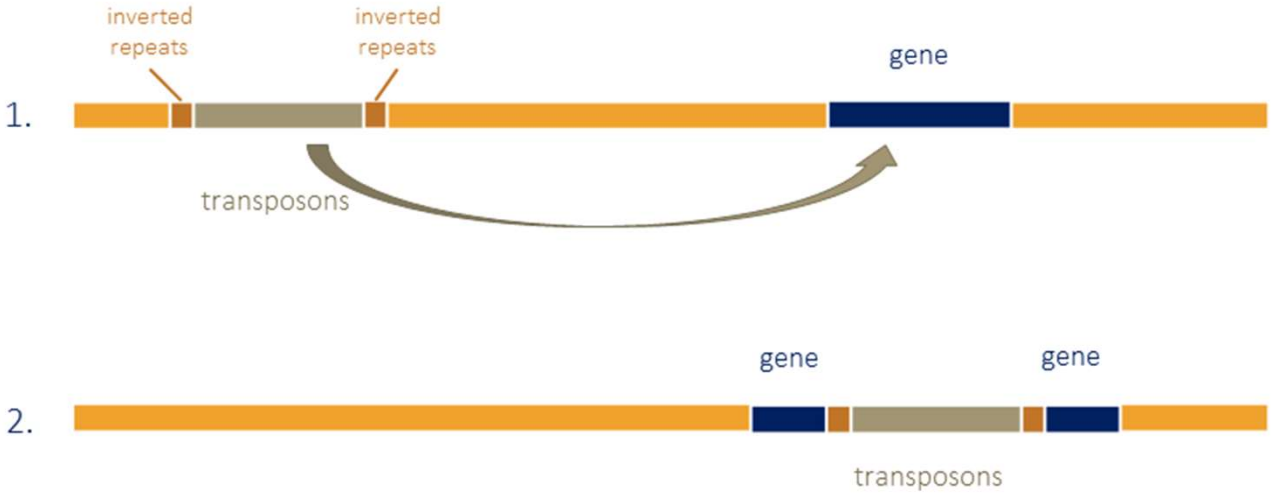


https://www.macmillanhighered.com/BrainHoney/Resource/6716/digital_first_content/trunk/test/hillis2e/hillis2e_ch15_6.html

MUNI SCI

Transposible / Mobile genetic elements

- DNA sequences that can move in the genome (randomly).
- They do not exist separately as plasmids or viruses.
- Defined by end sequences on both sites.
- Transposition can be both intramolecular and intermolecular.
- Significant source of genomic instability.



https://2014.igem.org/Team:Evry/Biology/Transposons

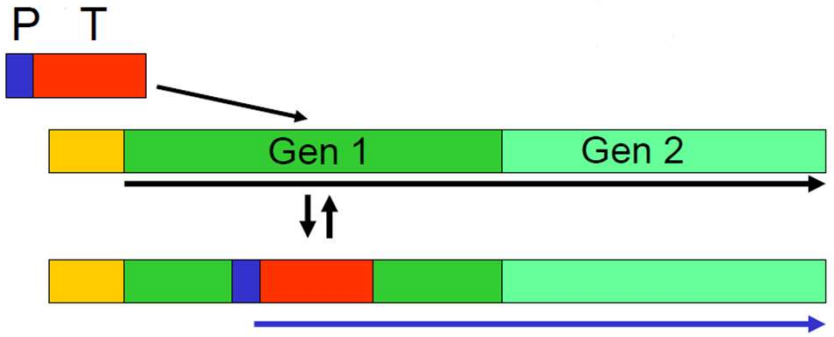
MUNI SCI

Transposible / Mobile genetic elements – transposons – jumping genes

- Widespread in eukaryotes and prokaryotes.
- 4% of the genome in yeast, 40% in humans, 70% in some amphibians and plants.
- Induce gene mutations (insertion inactivation, polar mutation).
- Induce changes in gene expression.
- Responsible for rearrangements of chromosomes or plasmids ('portable' sections of homology conditioning homologous recombination, interactions between genome components).
- Transfer of new traits (e.g. antibiotic resistance, oncogenes) between organisms (horizontal gene transfer).

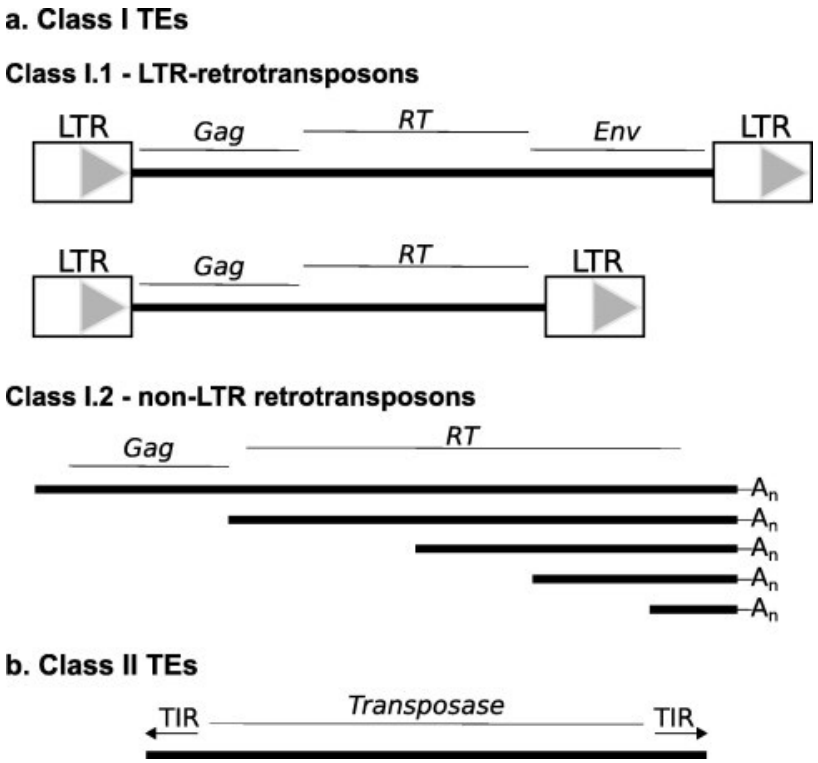
Specific characteristics of transposition

- The target sites are not homologous with the donor sites.
- Often accompanied by duplication of the transmitted sequence, i.e. the transposon remains in the original donor location.
- At the insertion site, short DNA sequences in the same direction are duplicated – transposon is at both ends surrounded by direct repetitions - a consequence of the transposition mechanism.
- Transposon insertion leads to activation, repression or modification of gene structure.
- After excision of transposition, the function is restored.



MUNI SCI

Transposable elements (TEs) - Mobile genetic elements



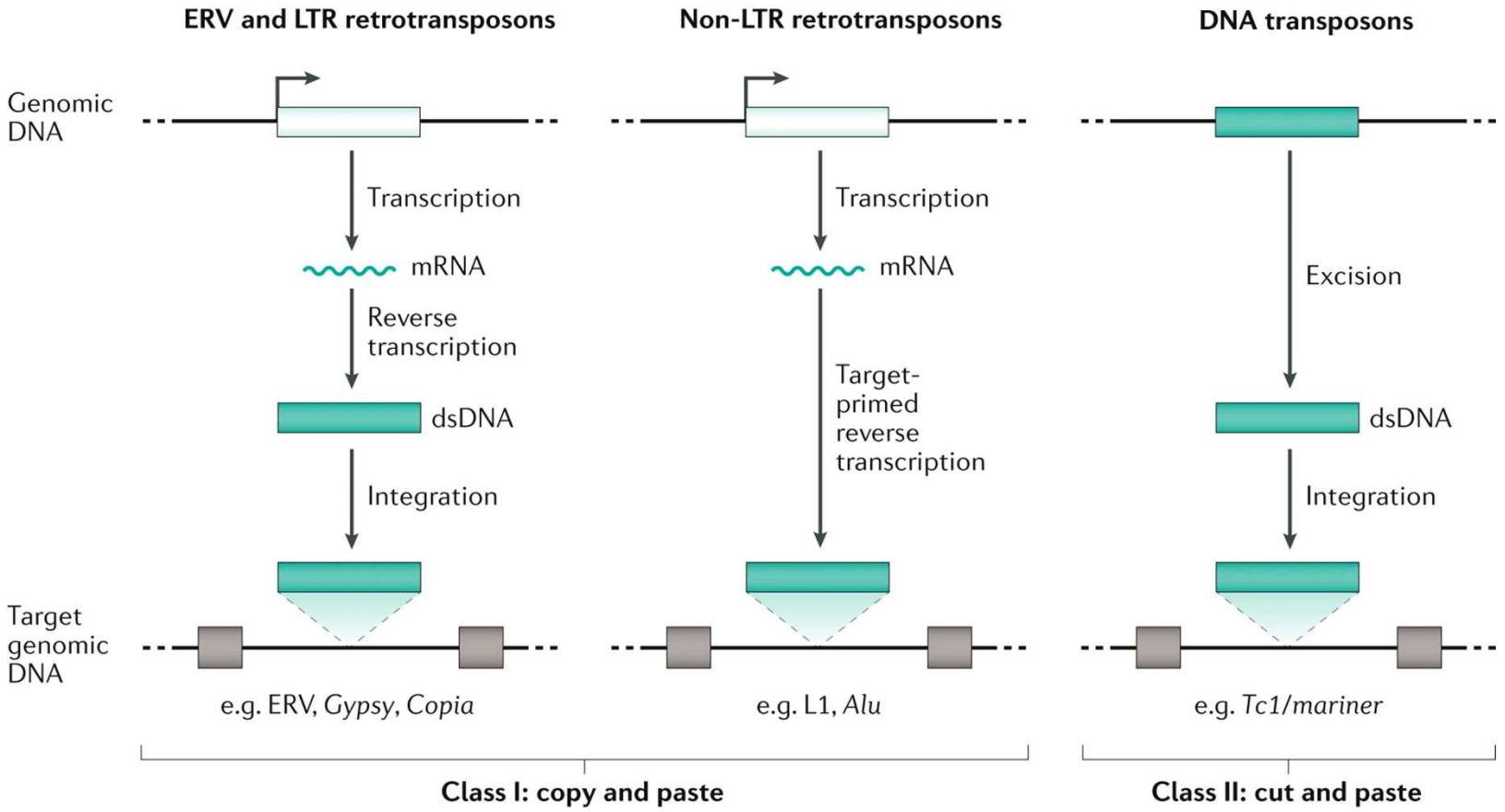
- Classification:
- Class I is composed of elements that transpose by reverse transcription of an RNA intermediate:
- Class I.1 TEs have all the signatures of an endogenous retrovirus-like element, including long terminal repeats at both ends and open reading frames (ORFs) coding for, a group antigen (Gag), a reverse transcriptase (RT), and in some case an envelope protein.
- Class I.2 elements only have Gag and RT ORFs and look like long retro-inserted messenger RNA (mRNA) with an A-rich tail at their 3′ end. Within a species of such elements many copies are truncated at their 5′ ends.
- Class II elements that transpose directly from DNA to DNA and have short terminal inverted repeats (arrowed) at both ends. They contain a gene coding a transposase, an enzyme required for their own transposition.

Molecular Phylogenetics and Evolution, Volume 86, May 2015, Pages 90-109
Lanciano S, et al. Nat Rev Genet. 2020. PMID: 32576954 Review.

MUNI SCI

Transposable elements (TEs) - Mobile genetic elements

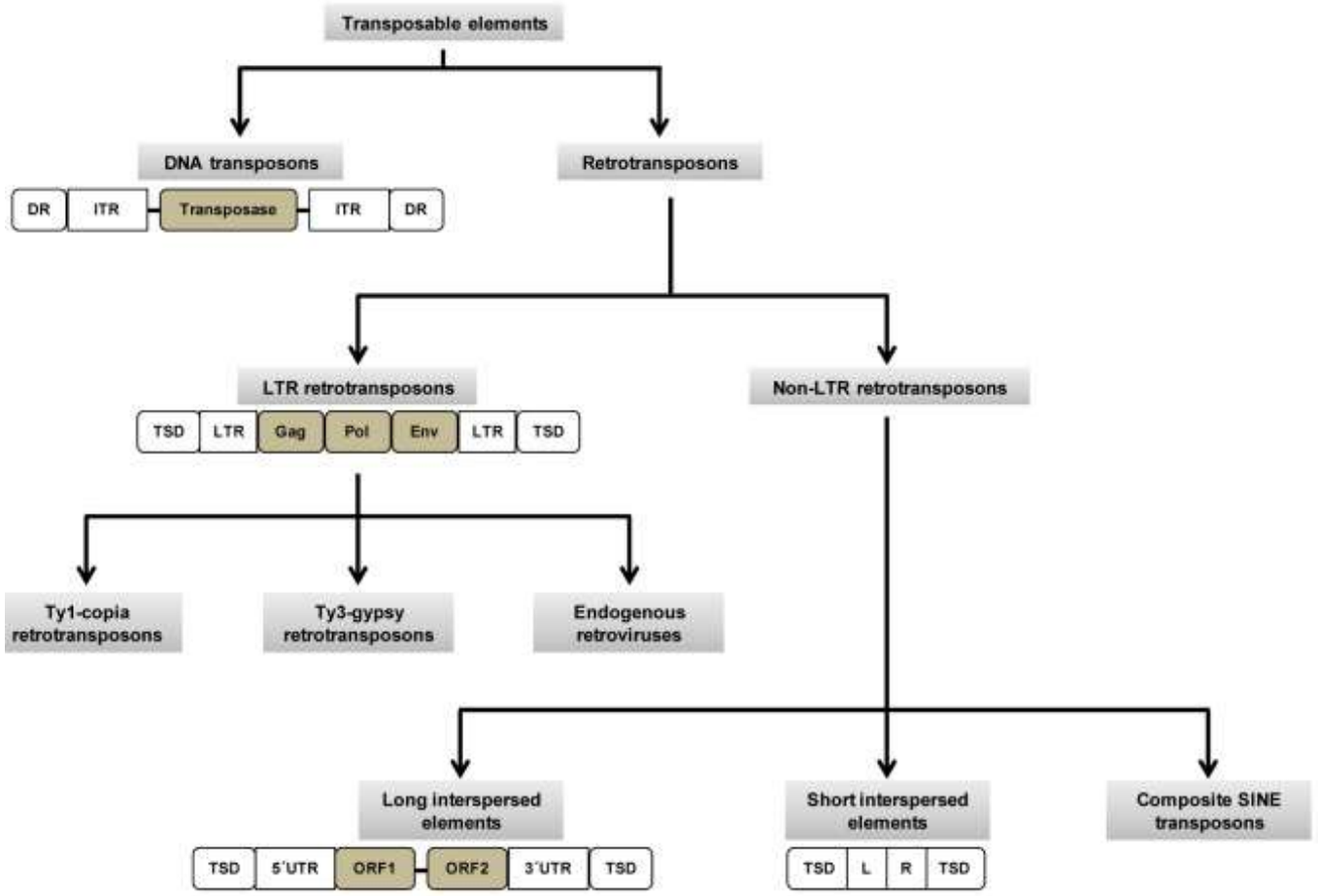
- Two main classes depending on their mobilization mechanism and molecular intermediates.



Lanciano S, et al. Nat Rev Genet. 2020. PMID: 32576954 Review.

MUNI SCI

Transposable elements (TEs)

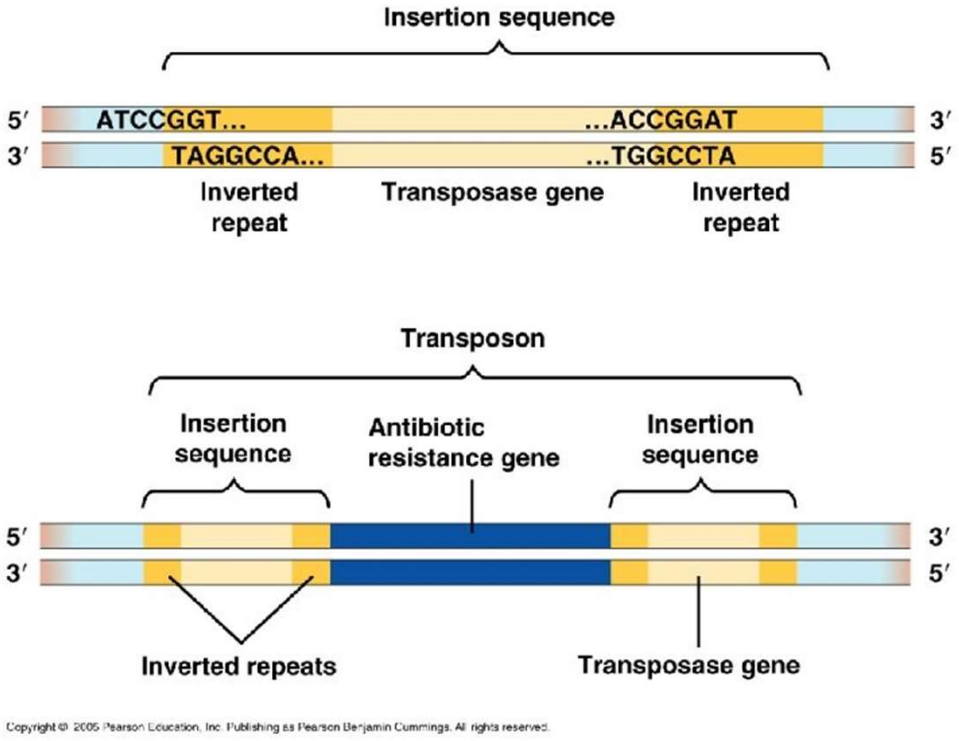


Genomics Inform. 2014 Sep; 12(3): 98–104.

MUNI SCI

Transposable elements (TEs) in prokaryotes

- Prokaryotes - two types of DNA Transposons:
- IS elements (Insertion sequences).
- Tn elements (Transposons).



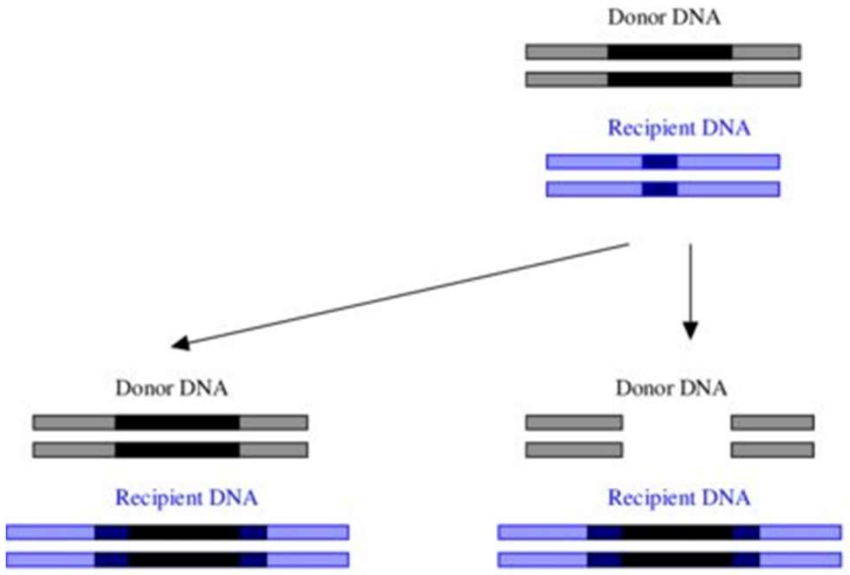
MUNI SCI

Transposable elements (TEs) in prokaryotes

- Tn elements (Transposons):

- Nonreplicative Tn „cut and paste“:
  - Set aside from the original location and integrated into the new one.
  - Key enzyme – transposase.
  - Prokaryotes and Eukaryotes.

- Replicative Tn „copy and paste“:
  - They are replicated during transposition (one copy remains in the original location, the other appears in the new location).
  - Key enzymes – transposase and resolvase.
  - Prokaryotes.



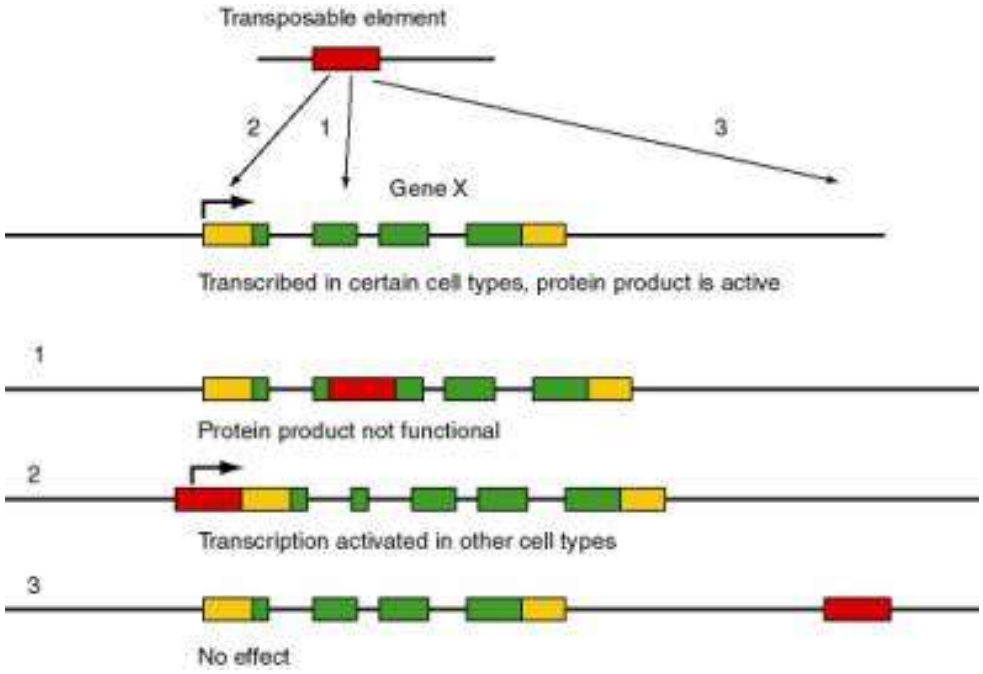
Replicative transposition

Nonreplicative transposition

MUNI SCI

Insertion sequence (IS) elements

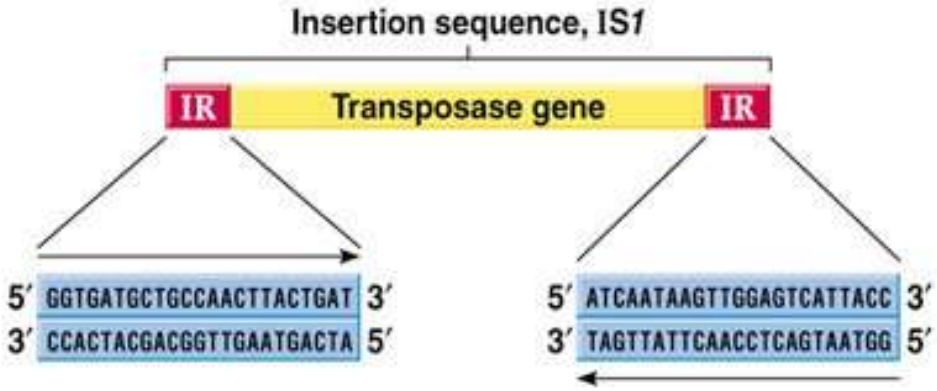
- Integration of an IS element may:
- Disrupt coding sequences or regulatory regions.
- Alter expression of nearby genes.
- Cause deletions and inversions in adjacent DNA.
- Result in crossing-over.



https://www.bx.psu.edu/~ross/workmg/TranspositionCh9.htm

Insertion sequence (IS) elements

- Simplest type of transposable element found in bacterial chromosomes and plasmids.
- Encode gene (transposase) for mobilization and insertion.
- Range in size from 768 bp to 5 kb.
- IS1 first identified in *E. coli's* galactose operon is 768 bp long and is present with 4-19 copies in the *E. coli* chromosome.
- Ends of all known IS elements show inverted terminal repeats (ITRs, IR).



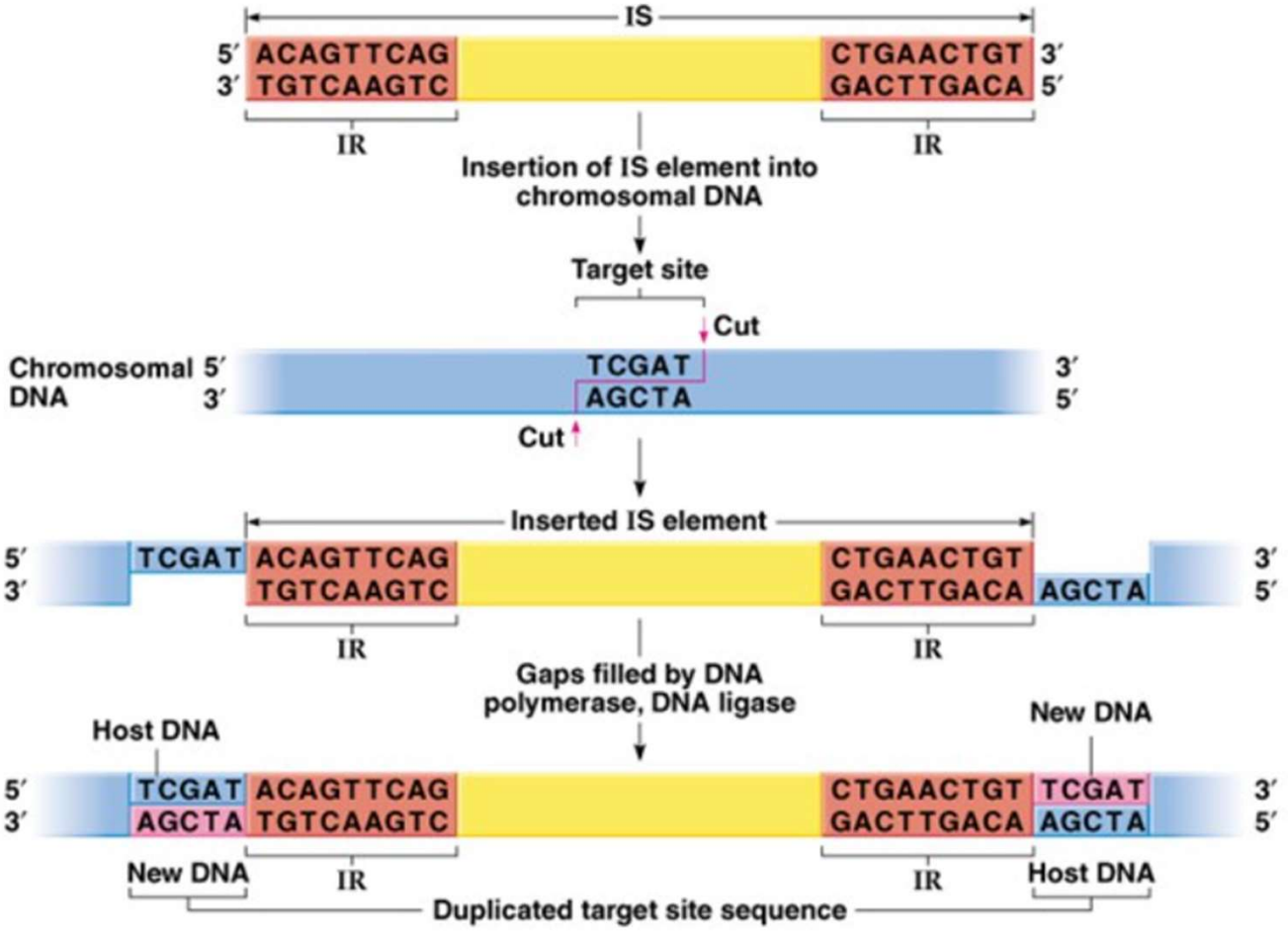
https://education.sakshi.com/en/csir-net/study-material/cell-biology/transposons-67119

MUNI SCI

Transposition of insertion sequence (IS) elements

- Original copy remains in place; new copy inserts randomly.
- Transposition requires transposase, coded by the IS element.
- IS element otherwise uses host enzymes for replication.
- Transposition initiates when transposase recognizes ITRs.
- Site of integration = target site.
- Staggered cuts are made in DNA at target site by transposase, IS element inserts, DNA polymerase and ligase fill the gaps (critically, transposase behaves like a restriction enzyme).
- Small direct repeats (~5 bp) flanking the target site are created.

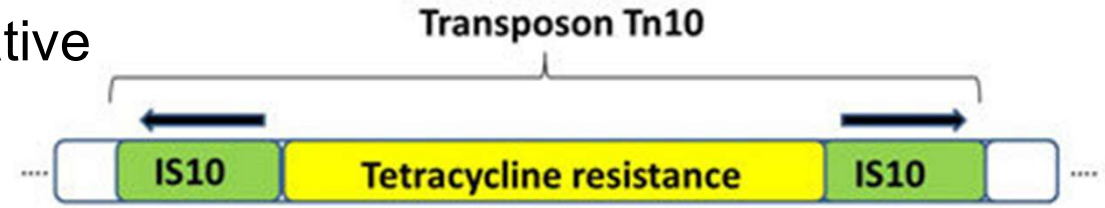
Transposition of insertion sequence (IS) elements



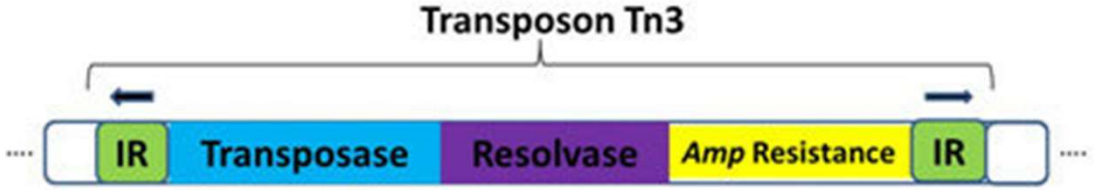
MUNI SCI

Transposons (Tn)

- Similar to IS elements but are more complex structurally and carry additional genes.
- Two types of transposons:
  - Nonreplicative



  - Replicative

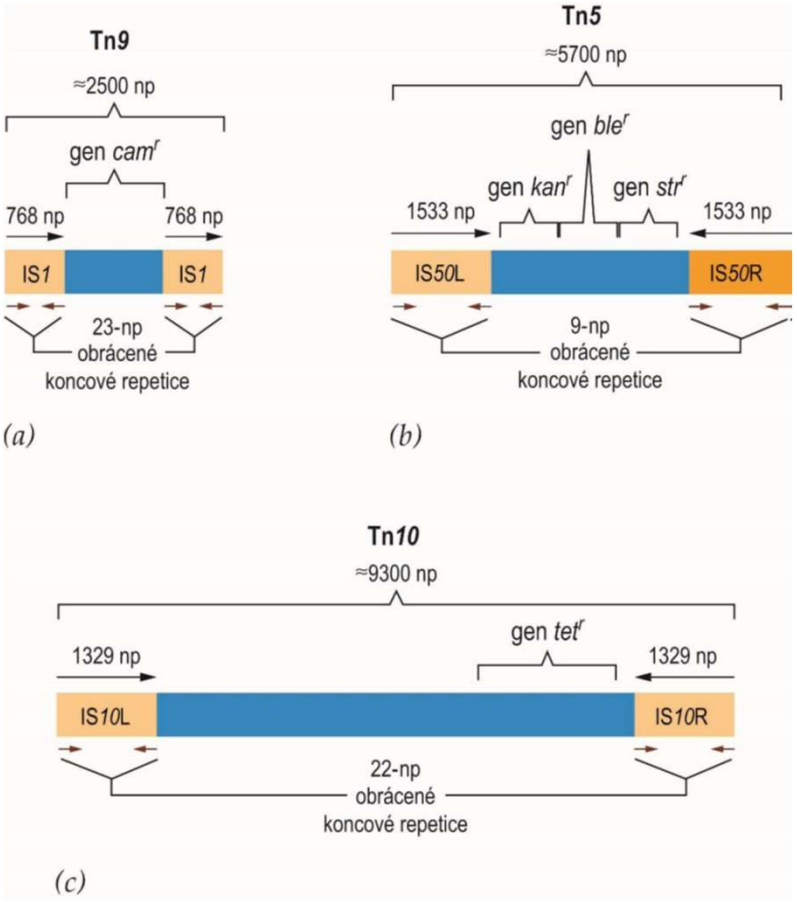


https://www.researchgate.net/publication/281264950_The_Dynamic_Genome_A_Gateway_of_Evolution

MUNI SCI

Assembled transposons

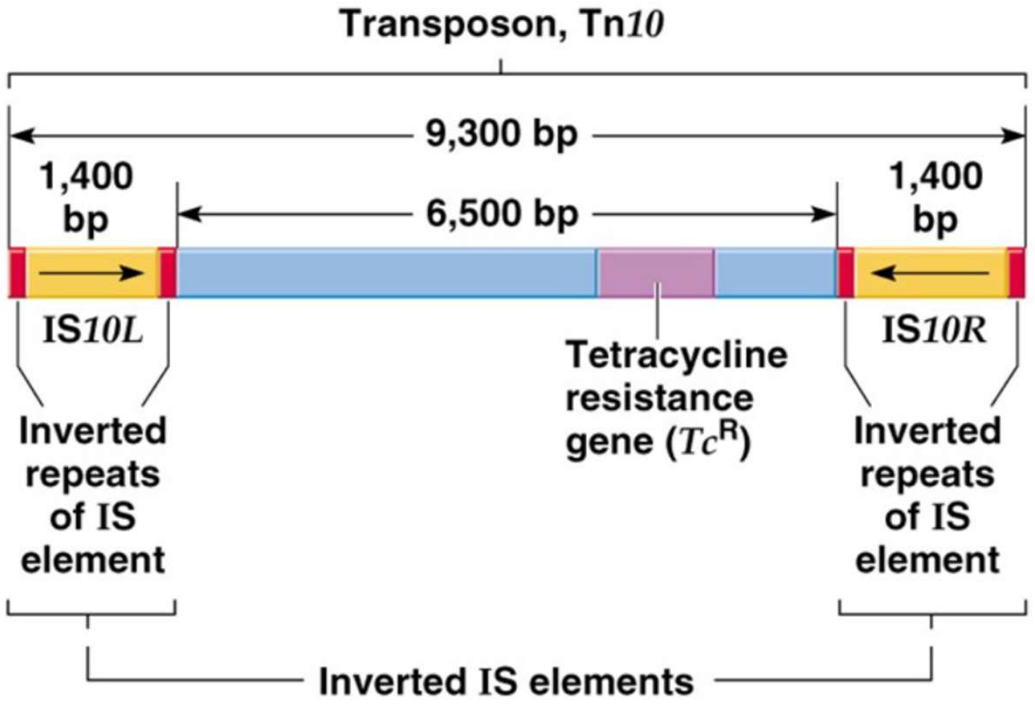
- Consequence of the occurrence of two IS elements in close proximity to each other.
- A pair of IS-elements will provide mobility to the intermediate region of DNA.
- End structures of IS-elements preserved.
  Carry e.g. Antibiotic resistance:
  - Kanamycin
  - Gentamycin
  - Ampicillin
  - Tetracycline
  - Chloramphenicol
  - Streptomycine



MUNI SCI

Composite transposons – Tn10

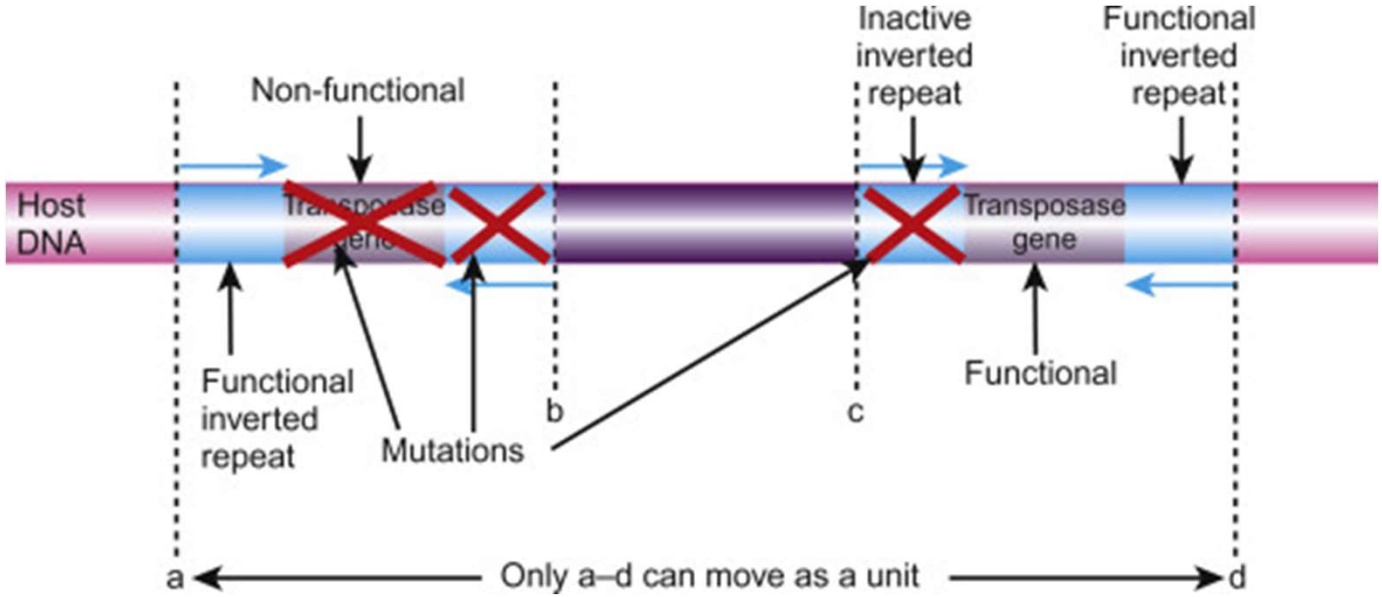
- Carry genes, for example gene for an antibiotic resistance, flanked on both sides by IS elements.
- Tn10 is 9.3 kb and includes:
  - 6.5 kb of central DNA (includes a gene for tetracycline resistance.
  - 1.4 kb inverted IS elements.
- IS elements supply transposase and ITR recognition signals.



MUNI SCI

Composite transposons – Tn10

- IS10 sequences at the ends are not identical.
- One IS encodes functional transposase.
- The other IS mutant (often the difference of a single nucleotide pair).

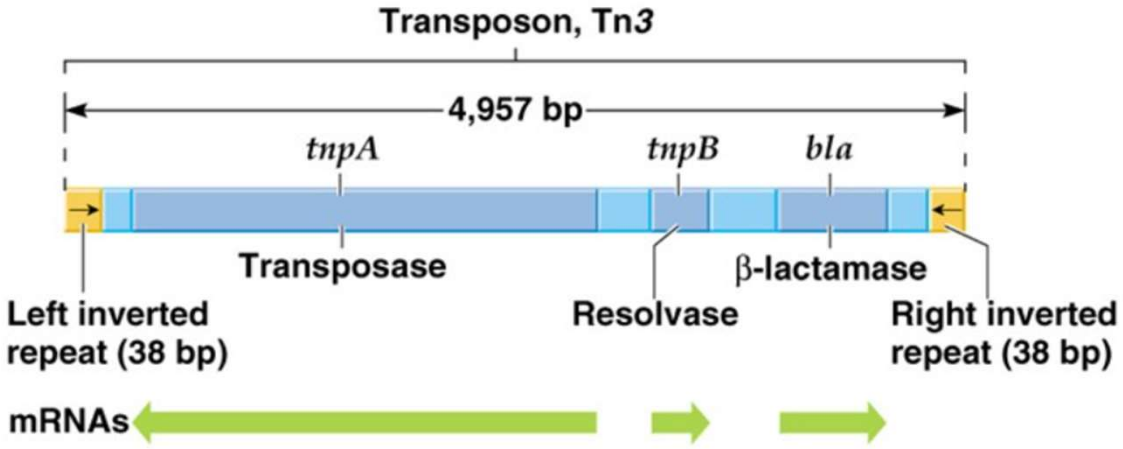


https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/composite-transposon

MUNI SCI

Replicative transposons – Tn3

- Carry genes, e.g. gene for antibiotic resistance, but do not terminate with IS elements.
- Ends are non-IS element repeated sequences only inverted repeats.
- It is a replicative transposon.



- Tn3 is 5 kb composed of:
- 38-bp ITRs
- 3 genes:
  - *bla* (β-lactamase)
  - *tnpA* (transposase)
  - *tnpB* (resolvase, which functions in recombination).

MUNI SCI

Transposition of transposons

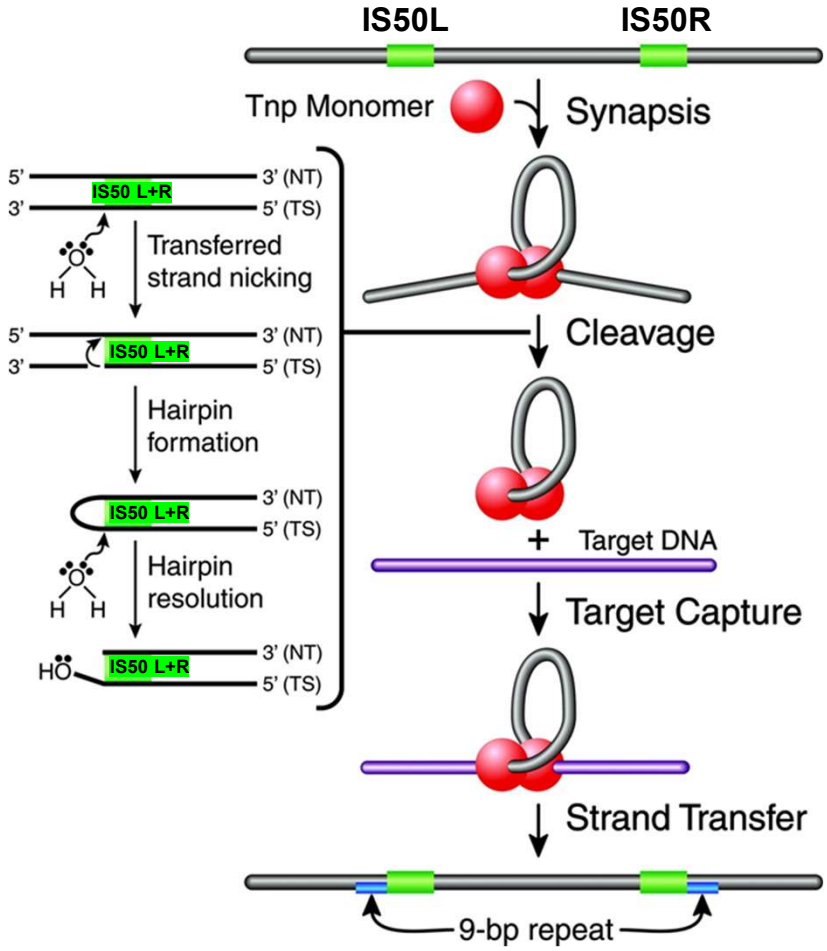
- All transposons use a common mechanism in which staggered nicks are made in target DNA, the transposon is joined to the protruding ends, and the gaps are filled.
- Similar to that of IS elements - duplication of short sequence at ends of target sites occurs.
- Cointegration = movement of a transposon from one genome (e.g., plasmid) to another (e.g., chromosome) integrates transposon to both genomes (duplication).
- Transposition replicative (duplication) or non-replicative (transposon lost from original site).
- Result in same types of mutations as IS elements: insertions, deletions, changes in gene expression, or duplication.
- Crossing-over occurs when donor DNA with transposable element fuses with recipient DNA.

Non-replicative transposon

- Non-Replicative transposon leaves its original place and move to the another location in the genome - “Cut and Paste”.
- This type of mechanism requires only a transposase.
- The insertion elements and composite transposons like Tn5 and Tn10 use this mechanism.
- Non-replicative transposons leave a break in the donor molecule which is lethal to the cell unless it is repaired.

Mechanism of non-replicative transposon transposition

- Transposon 5 - Tn5.
- Transposition initiated by Transposase (Tnp) binding to the transposon specific Insertion Sequences (IS50), and the formation of a synaptic complex (SC) by a process called synapsis.
- The SC contains Transposase dimer and two IS50 L+R.
- Catalytic cleavage activated H_2O coordinated by Mg^{2+} nicks the transferred DNA strand on both sides by a nucleophilic attack, forming a 3′- hydroxyl group.
- The free 3′-hydroyxl group acts as a nucleophile and cleaves the non-transferred DNA strand (NT), forming a hairpin.

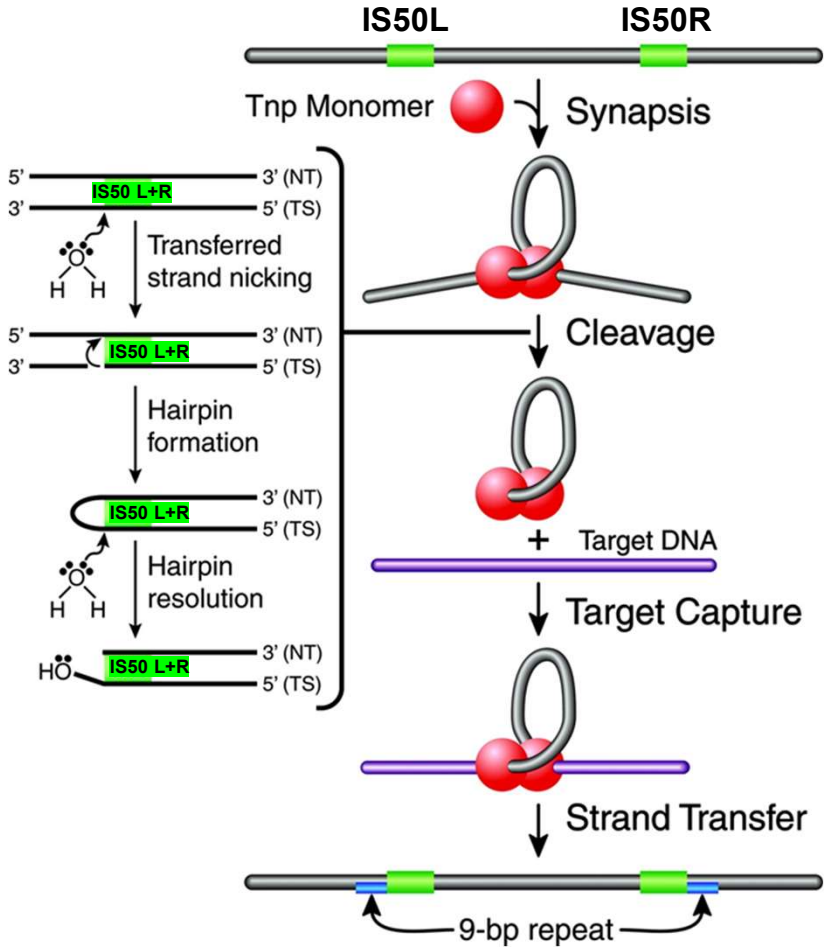


Journal of Bacteriology, Volume 190, Issue 4, 15 February 2008, Pages 1484-1487.

MUNI SCI

Mechanism of non-replicative transposon transposition

- A second activated water molecule resolves the hairpin, resulting in a double-stranded DNA cleavage product.
- The post-cleavage synaptic complex is now free to bind to target DNA through target capture. The 3′-hydroxyl group of the transposon end attacks the phosphodiester backbone of target DNA during
- strand transfer.
- A 9-bp duplication in the target results, due to the staggered strand transfer reactions followed by DNA repair by host enzymes.



Journal of Bacteriology, Volume 190, Issue 4, 15 February 2008, Pages 1484-1487.

MUNI SCI

Replicative transposon

- Replicative transposon is first replicated and then one of the copy will move to the another location in the genome. Thus, the transposon will remain on its original position - "Copy and Paste".

- Replicative transposition involves two types of enzymatic activity:
  - Transposase that acts on the ends of the original transposon.
  - Resolvase that acts on the duplicated copies.

- Replicative transposition occurs through a cointegrate formation, which is produced by fusion of two replicons, one originally possessing a transposon, the other lacking it; the co-integrate has copies of the transposon present at both junctions of the replicons.

- Resolution occurs by a homologous recombination mediated by resolvase enzyme between the two copies of the transposon in a co-integrate leading to the donor and target replicons, each with a copy of the transposon.

Mechanism of replicative transposon transposition

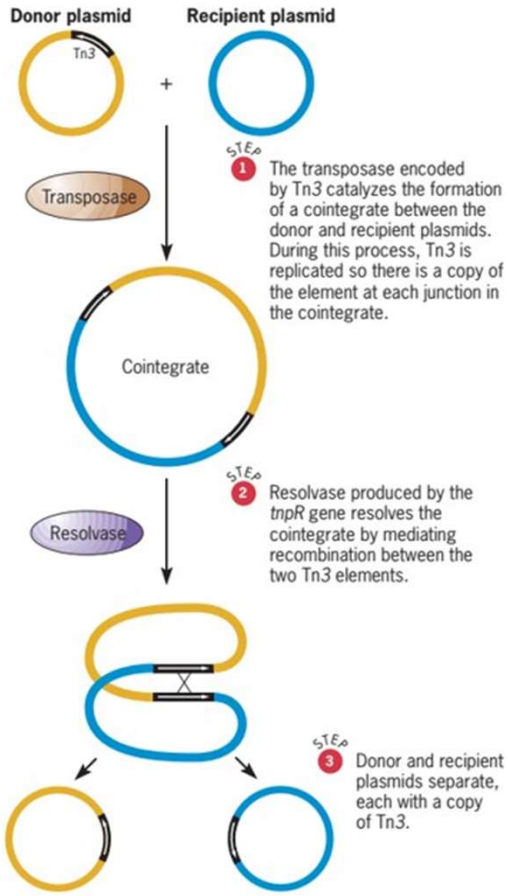


FIGURE 17.6 Transposition of Tn3 via the formation of a cointegrate.

- Non-composite transposons (Tn3) is a replicative transposons that undergoes transposition in two stage process.
- In the first stage, two plasmid - one containing Tn3 transposons; donor plasmid, and the other recipient plasmid undergoes fusion catalyzed by transposase enzymes giving rise to a structure called cointegrate.

https://www.onlinebiologynotes.com/transposable-elements-characteristics-and-mechanisms-of-transposition/

MUNI SCI

Mechanism of replicative transposon transposition

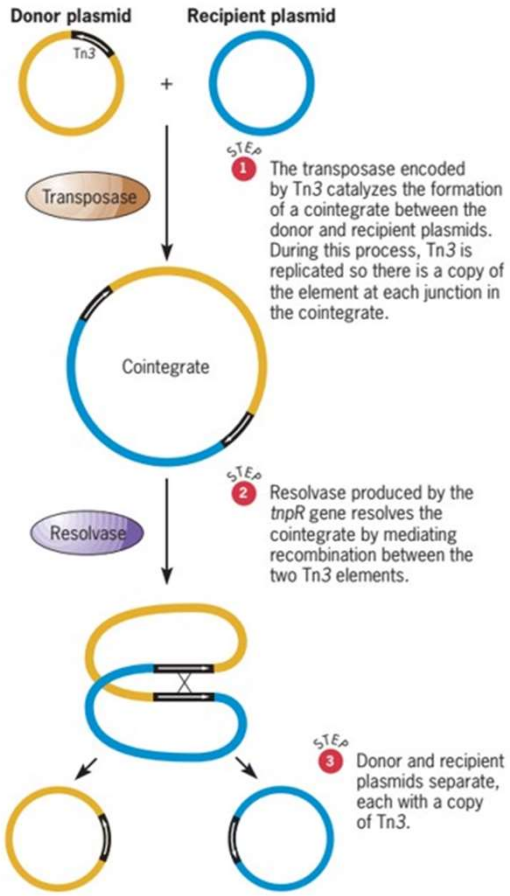


FIGURE 17.6 Transposition of Tn3 via the formation of a cointegrate.

- During the formation cointegrate, Tn3 is replicated, and one copy of Tn3 is inserted at each fusion point, where the two plasmids have fused.
- In the second stage of transposition, the resolvase mediates a site-specific recombination between the two Tn3 copies at the resolution site, and when it is completed, cointegrate is resolved into two plasmids, each with a copy of Tn3.

https://www.onlinebiologynotes.com/transposable-elements-characteristics-and-mechanisms-of-transposition/

Transpoable elements (TEs) in eukaryotes

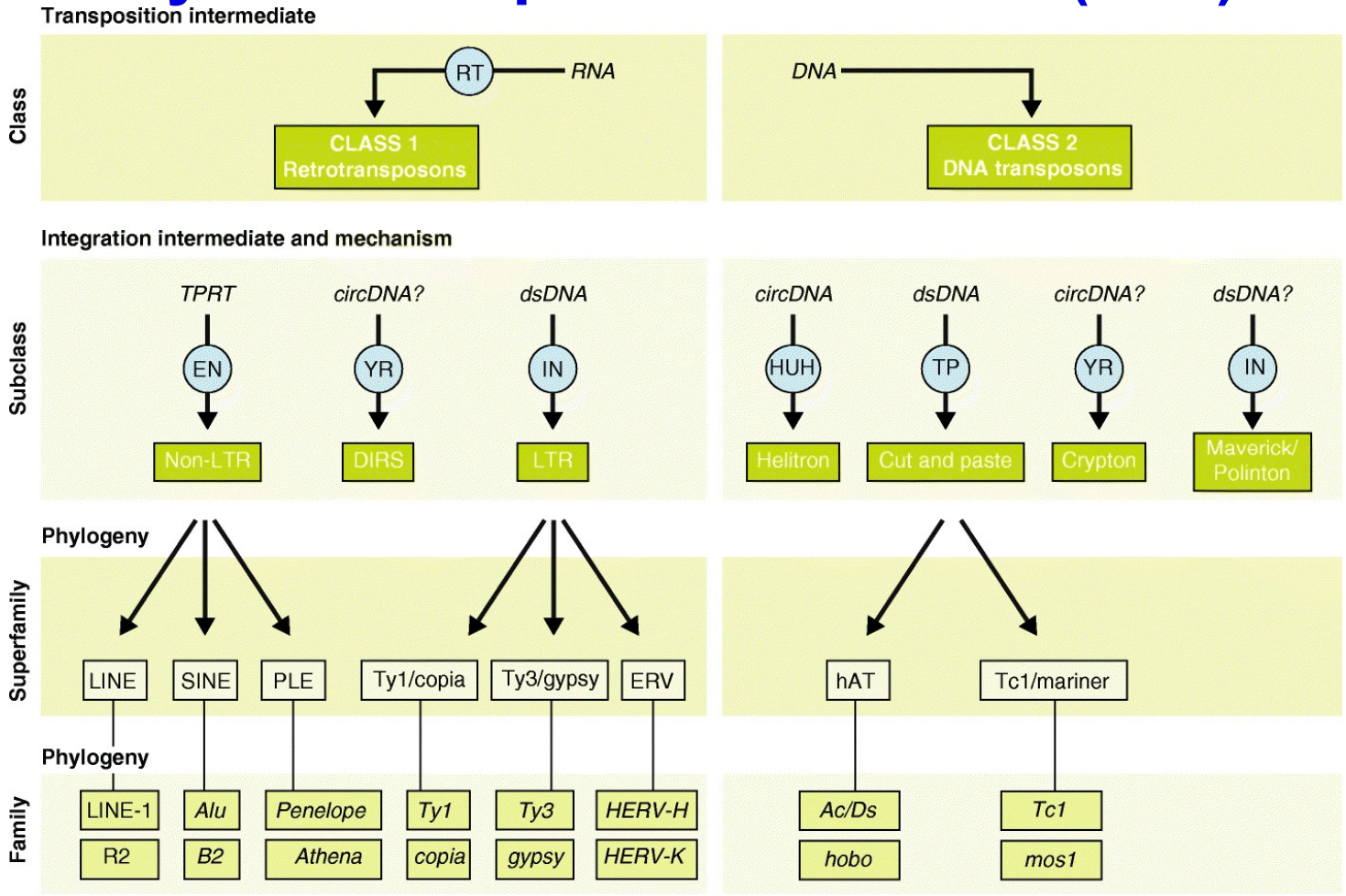
- Eukaryotic genomes have many copies of transposons (~45% of the human genome).

- Transposition accompanies insertion in a genome site and some (not all) insertions can cause changes in gene expression, its regulation and products and can lead to speciation, evolvement of new distinct species during evolution, or diseases.

- The insertion sites are random, although some sites, called hot spots, are preferred to others.

Classification eukaryotic transpoable elements (TEs)

- Schematic and examples showing the key features and relationships between TE classes, subclasses, superfamilies, and families.



Genome Biology, 2018, volume 19, Article number: 199.

MUNI SCI

General properties of plant DNA transposons

- Possess ITR sequences and generate short repeats at target sites.

- May activate or repress target genes, cause chromosome mutations, and disrupt genes.

- Two types:
- Autonomous elements transpose themselves; possess transposition gene.
- Nonautonomous elements do not transpose themselves; lack transposition gene and rely on presence of another Tn - transposon.

- McClintock demonstrated purple spots in otherwise white corn (*Zea mays*) kernels are results of both these types of transposable elements.

General properties of plant DNA transposons

MUNI SCI

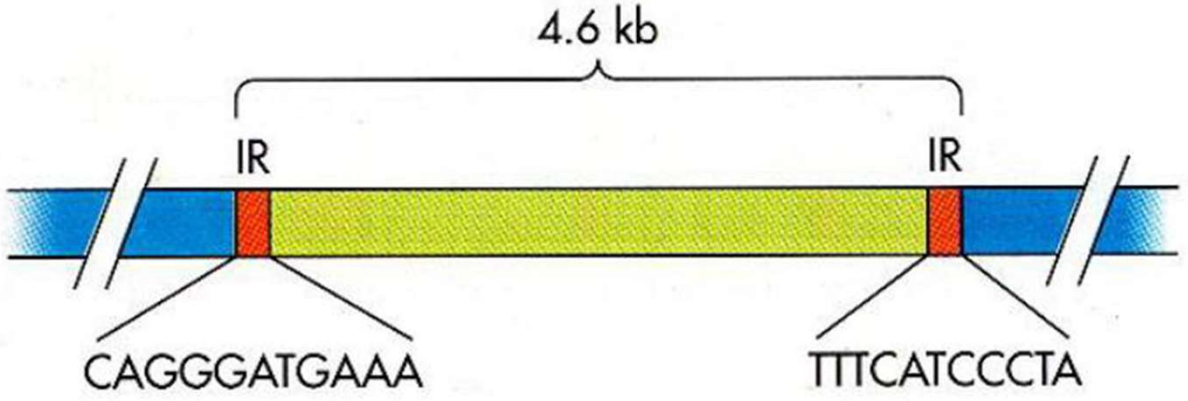
McClintock´s discovery of transposon in corn

- c/c = white kernels and C/- = purple kernels.
- Kernel color alleles/traits are “unstable”.
- If reversion of c to C occurs in a cell, cell will produce purple pigment and a spot.
- Earlier in development reversion occurs, the larger the spot.
- McClintock concluded “c” allele results from a non-autonomous transposon called “Ds” inserted into the “C” gene (Ds = dissociation).
- Autonomous transposon “Ac” (Ac = activator) controls “Ds” transposon.

https://www.slideshare.net/DeepakKumarGupta2/general-aspect-of-neoplasia

McClintock´s discovery of transposon in corn

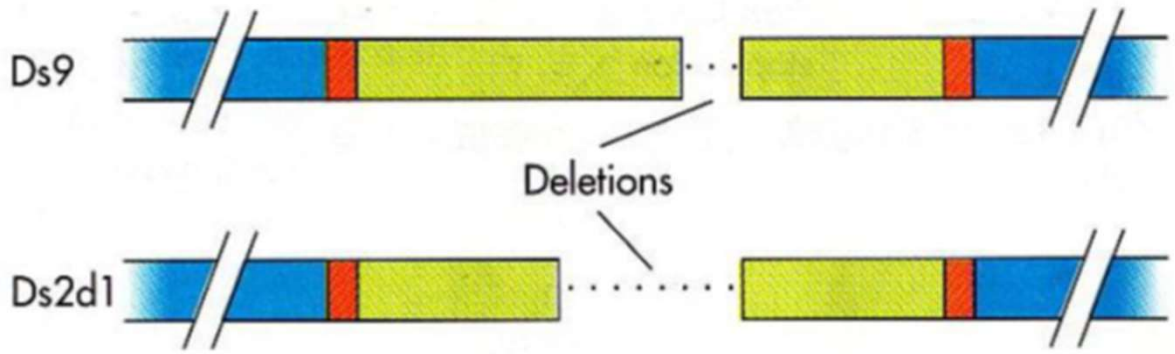
- Ac-element –activator:
- Consisting of 4563 bp restricted by inverted repeats 11 bp long and 8 bp long straight repeats (they are formed by duplication at the target point and are not an integral part of the element).
- Contains a gene for transposase.
- Ac element is autonomous.



MUNI SCI

McClintock´s discovery of transposon in corn

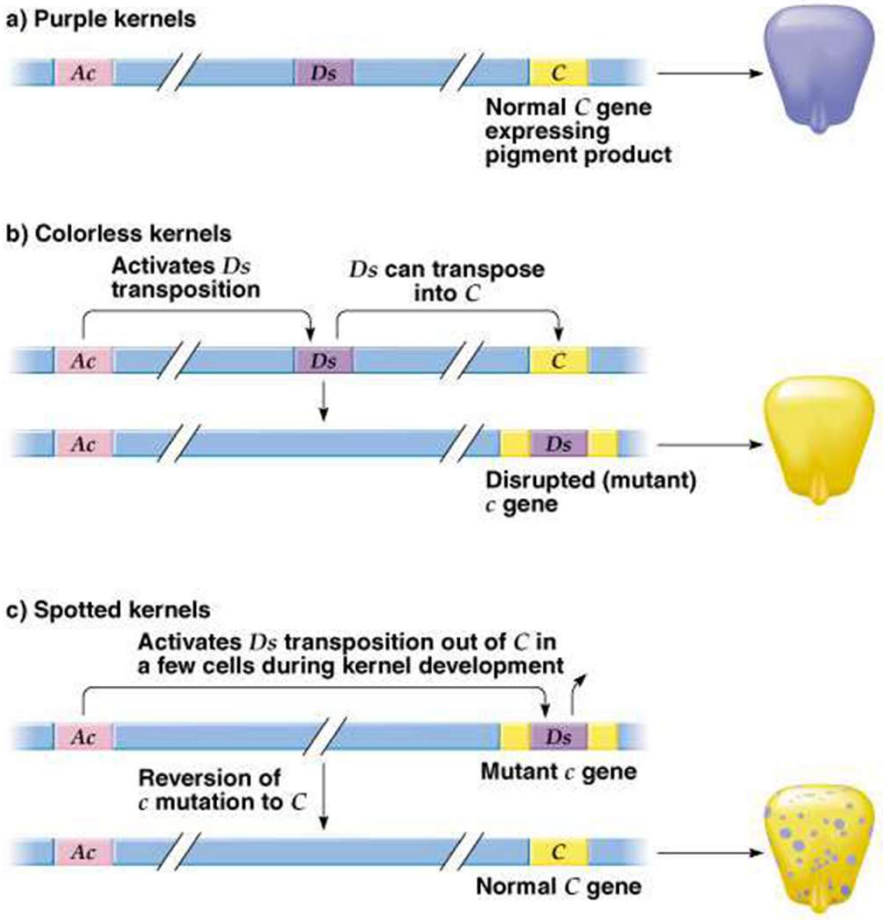
- Ds-elements –dissociator:
- Structurally heterogeneous - have the same terminal repeats as Ac-elements, internal sequences are different.
- For their transposition they need a transposase of an Ac-element.
- This transposase is therefore a transacting protein.
- Ds element is nonautonomous.



MUNI SCI

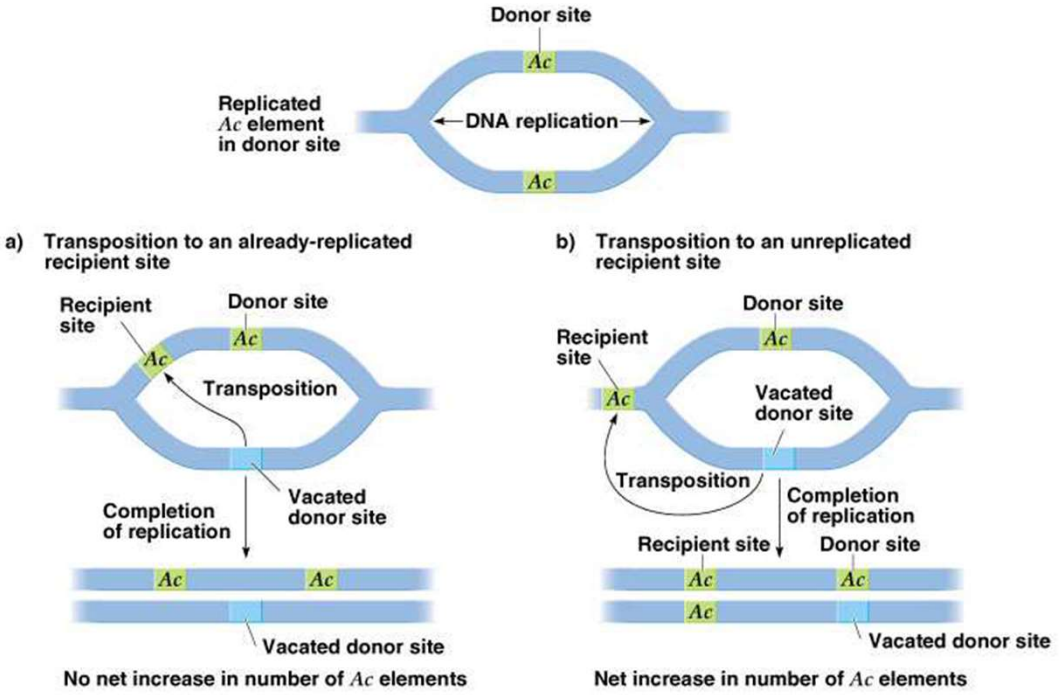
Transposon effect on corn kernel color

- A corn plant with a genotype of c/c will have colorless kernels, while a genotype of C/- will have purple ones (C gene produces the purple pigment)
- The Ac-Ds transposon system controls the distribution of color in a kernel. Ds (dissociation) elements are nonautonomous. Ac (activator) elements are autonomous.
- If a reversion occurs in a cell (Ds is removed from the mutant c gene), that cell will produce a purple pigment (c → C) . In the case of the figure, the reversion appears to be late, so the kernel is mostly colorless.
- Ac/Ds are developmentally regulated. Ac transposes is active only during chromosome replication.



MUNI SCI

Transposon effect on corn kernel color



- Ac transposes is active only during chromosome replication, employing a conservative mechanism.
- Two possible results of Ac transposition, depending on whether the target DNA has replicated.
- A) If transposition takes place into an already replicated recipient site, there is no increase in the number of Ac elements.
- B) If transposition takes place into an unreplicated recipient site, there is a net increase in the number of Ac elements.
- Ds replication is the same (but driven by an Ac element).

MUNI SCI

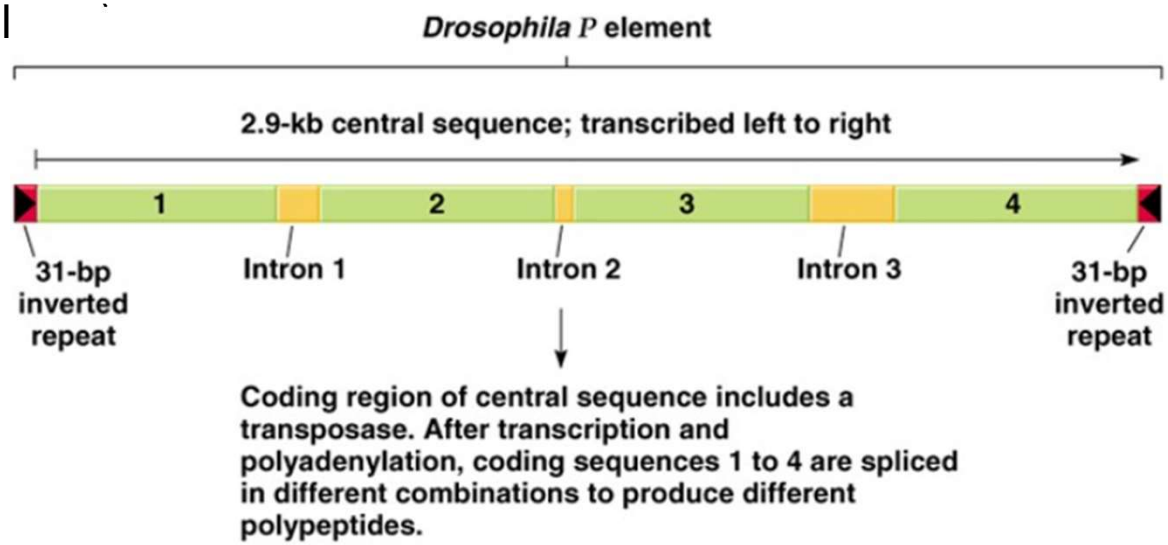
Transposable element in *Drosophila*

- It is estimated that 15% of Drosophila genome is mobile.
- Crossing certain strains of drosophila produces hybrids characterized by a set of aberrant characteristics, including numerous mutations, chromosome breaks and sterility.
- This syndrome of abnormalities was named dysgenesis of hybrids (from Greek "deterioration of quality").
- Strains of drosophila can be divided into two types – M and P - depending on whether or not their crossing leads to dysgenic hybrids.
- Only the crossing between M and P strains leads to the emergence of dysgenic species, where the male comes from the P strain, the female from the M strain.

Transposable element in *Drosophila*

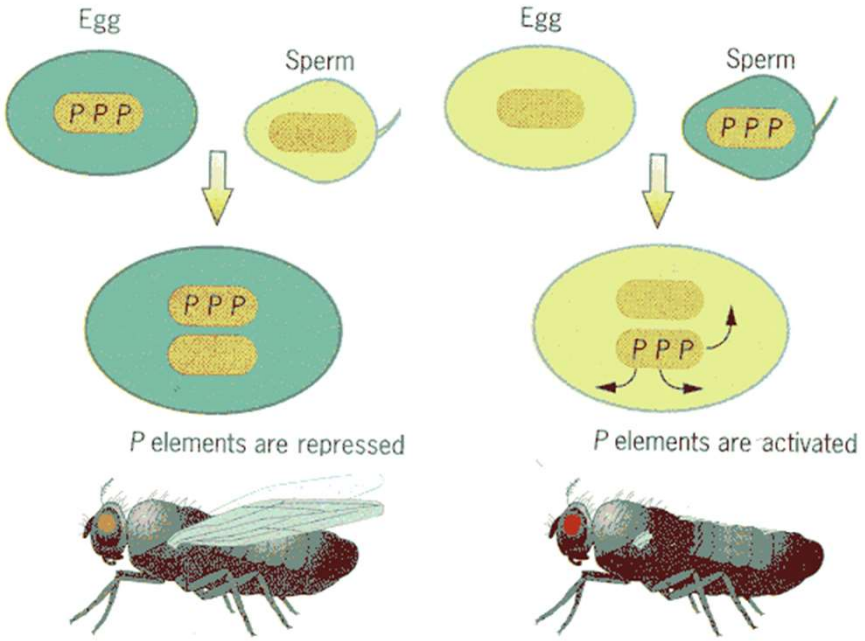
- Dysgenesis of hybrids is mediated by presence of P element.

- P-elements
- Are transposable elements that carry genes for transposase activity that cause the elements to move, and repressor piwi-interacting RNA activity that prevents expression of transposase.
- P elements vary in length from 500 - 2,900 bp.
- P strains code a repressor „ piwi-inteacting RNA“ present in the cytoplasm, which makes P elements stable in the P strains (but unstable when crossed to the wild type female; female lacks repressor in cytopl



MUNI SCI

P-element-mediated hybrid dysgenesis in *Drosophila*



- Hybrid dysgenesis occurs when haploid genome of male (P strain) possesses ~40 P elements/genome.
- In a cross between a P-element-carrying female and a laboratory male, repressors in the maternally - derived cytoplasm repress expression of the maternally - inherited P elements. The resulting offspring show the wild-type phenotype.
- In a cross between a P-element-carrying male and a laboratory female, repressors are absent in the maternally - derived cytoplasm. The two zygotes are chromosomally identical but cytoplasmically different. In the right-hand cross, P elements are activated and undergo transposition in the genome, causing release of mutator activity and a variety of dysgenic phenotypes in the offspring.

https://www.mun.ca/biology/scarr/P-element_hybrid_dysgenesis.htm

Repression of hybrid dysgenesis in *Drosophila*

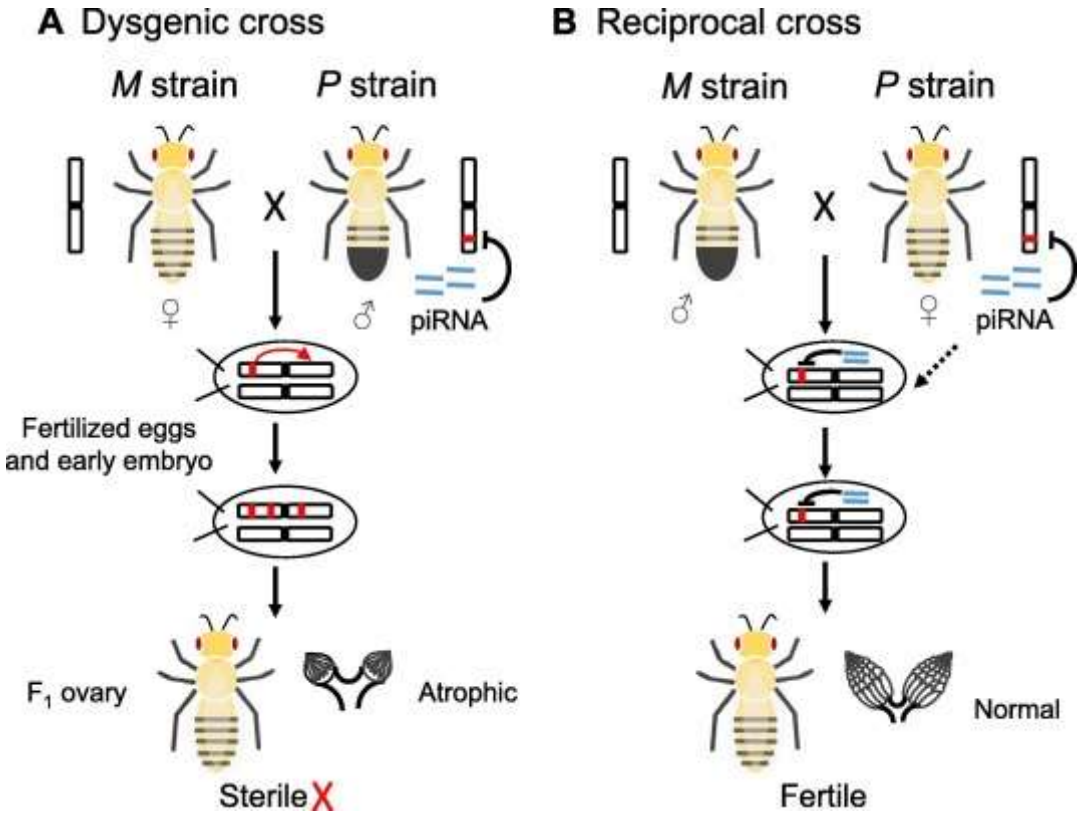
- Drosophila can prevent the effects of P-elements by RNA-interference using piRNA (derived from the P-elements themselves).
- piRNA (PIWI-interacting RNA) is a type of small RNA that can provide targeted degradation of the mRNA form the P element.
- piRNA (length 26-31 nts) form complexes with a specific group of proteins called "piwiproteins" (P-Element Induced Wimpy Testis).
- Female P-strains form piRNA and pass them on to their offspring through cytoplasm.
- piRNA restricts P-element activity in the embryonic line.
- Preventing hybrid dysgenesis.
- Maternal transmission of repressing piRNA explains why the offspring of crossing P-females with M-males and P-females with P-males is not dysgenic.

Repression of hybrid dysgenesis in *Drosophila*

- piRNA prevents hybrid dysgenesis only if it is present in the cytoplasm of the egg.

- A. Dysgenic cross: the crossing between M females (without P element) and P males (with P element) produces sterile offspring since the active transposition of P element disrupts genome and induces gonadal atrophy.

- B. Reciprocal cross: the crossing between M males and P females produces fertile offspring since the maternally inherited piRNAs repress activities of P elements in the offspring.



Genomics, Proteomics & Bioinformatics, Volume 15, Issue 3, June 2017, Pages 164-176.

Retroelements

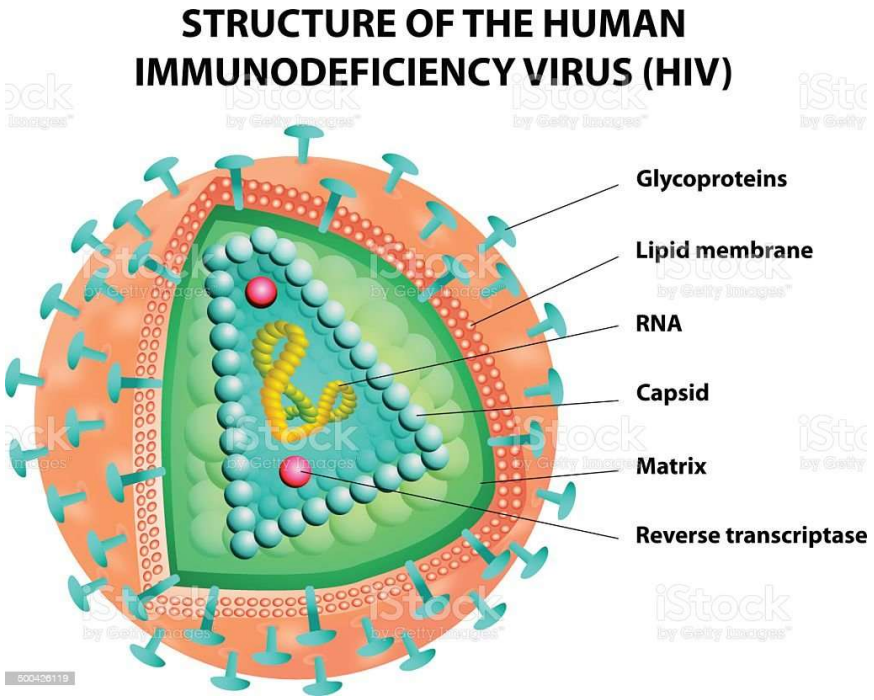
- Require reverse transcription of RNA into DNA → this reversal flow of genetic information led to the definition "retro", that is, "reverse" (lat.).
- Retroviruses - their genome consists of single-stranded RNA, which is converted into double-stranded DNA after infection of the host cell, with the participation of in reverse transcriptase they are able to leave the cell.
- Retrotransposons - limited to their own genome.
  - Elements similar to form DNA from RNA by reverse transcription retroviruses.
  - Mobile genetic elements which DNA is formed by reverse transcription of polyadenylated RNA.

Retroviruses

- RNA viruses.
- Discovered in connection with the development of certain tumors in chickens, cats and mice (Peyton Rous; 1966 NC).
- In 1970: Discovery of reverse transcriptase (David Baltimore, Howard Temin; 1975 NC).

- HIV (Human Immunodeficiency Virus) causing AIDS in humans (Acquired Immune Deficiency Syndrome).

- Prototypical retrovirus, life cycle and genome structure studied in detail.



https://www.istockphoto.com/cs/vektor/diagram-viru-hiv-vektor-gm500426119-42895954

MUNI SCI

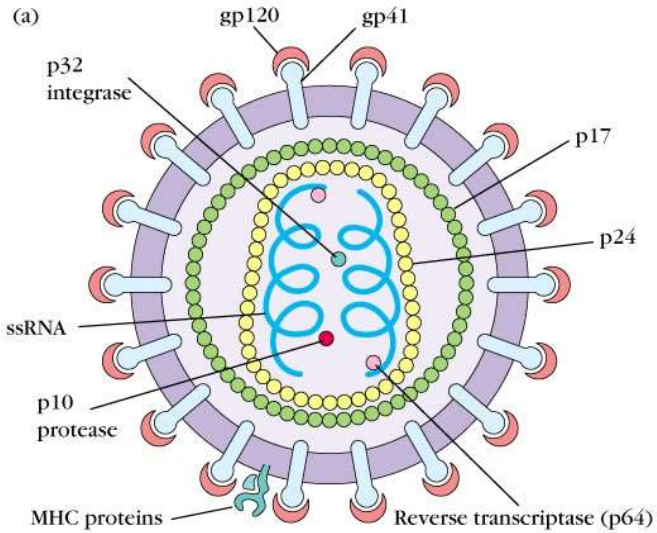
HIV genome and structure

- Formed by two identical single-strand RNA molecules; inside the viral particle is found along with several proteins including two reverse transcriptase molecules.

- Slightly larger than 10 kb.

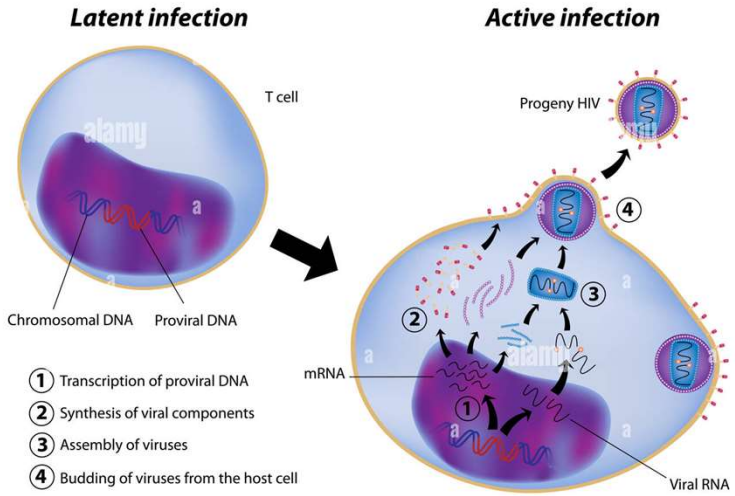
- Contains several genes:
- Gag - viral particle proteins
- Pol - reverse transcriptase and integrase
- Env - glycoproteins of the viral lipid envelope (gp120; gene for virulence)

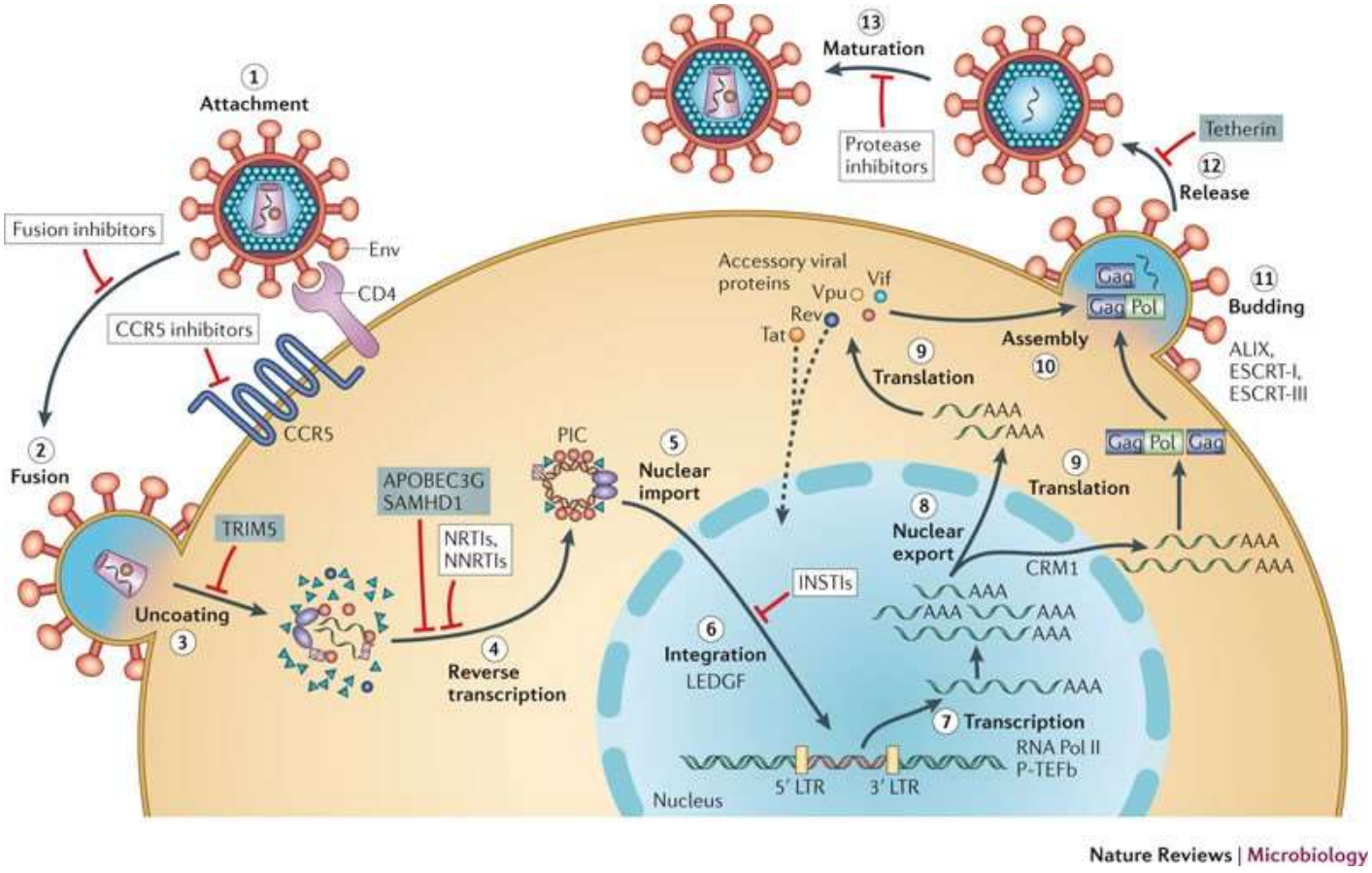
- Reverse transcriptase converts a single-stranded RNA into a double-stranded DNA and it is randomly incorporated into the chromosome of infected cells, which contains many copies of viral genomes.



https://biosci.mcdb.ucsb.edu/immunology/Immunodeficiencies/HIV-structure.htm

HIV replicative cycle





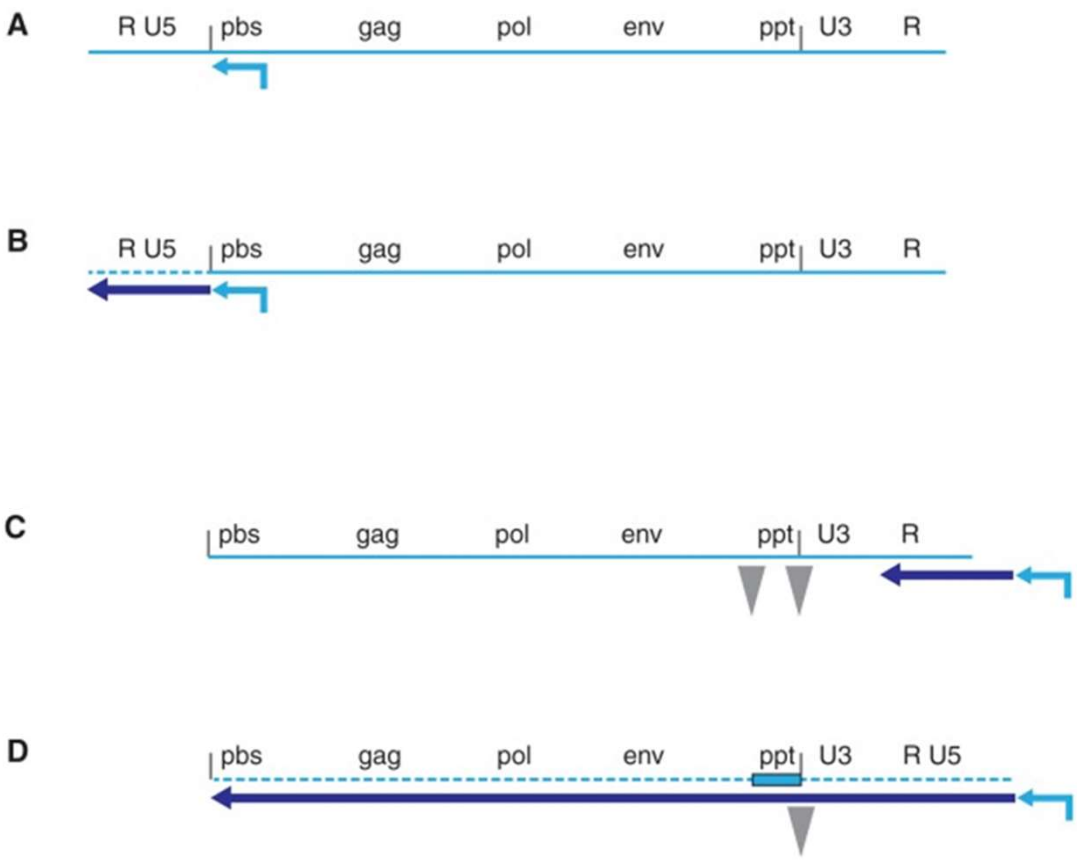
https://www.alamy.com/stock-photo-latent-and-active-infection-by-hiv-49485818.html

Nature Reviews Microbiology, 2012, Volume 10, pages279–290.

MUNI SCI

HIV reverse transcription

- (A) The RNA genome of a retrovirus (light blue) with a tRNA primer base paired near the 5′ end to primer binding site - pbs sequence.
- (B) RT has initiated reverse transcription, generating minus-strand DNA (dark blue), and the RNase H activity of RT has degraded the RNA template (dashed line).
- (C) Minus-strand transfer has occurred between the R sequences at both ends of the genome, allowing minus-strand DNA synthesis to continue.
- (D) RNA degradation, A purine-rich sequence (ppt), adjacent to U3, is resistant to RNase H cleavage and serves as the primer for the synthesis of plus-strand DNA.

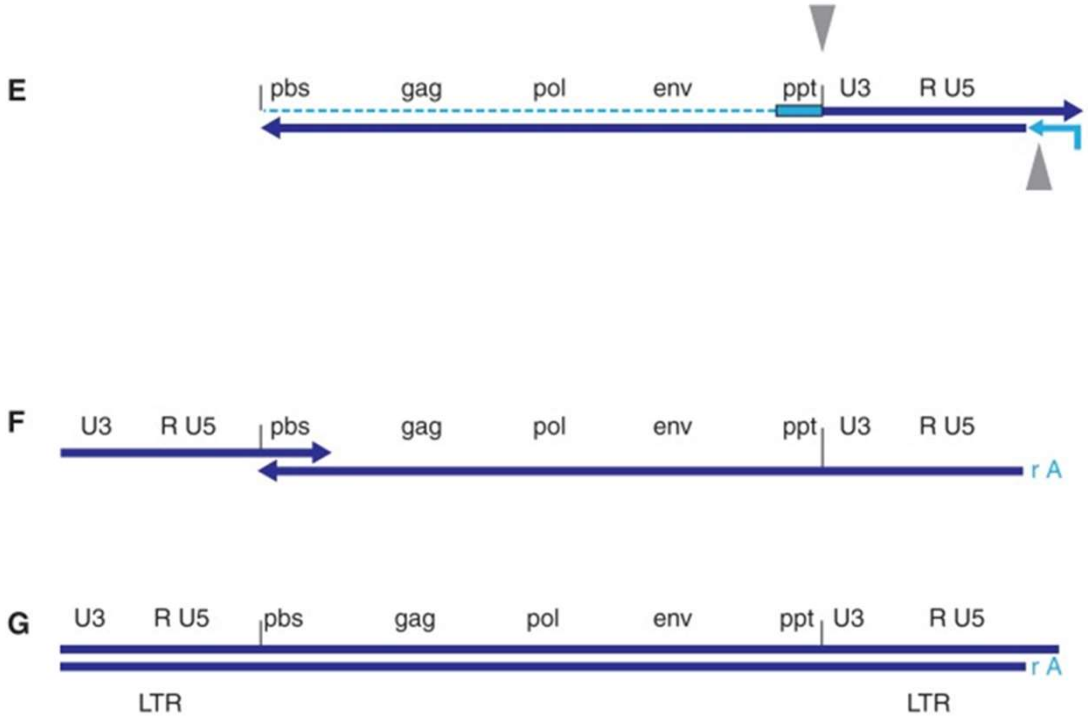


Cold Spring Harb Perspect Med. 2012 Oct; 2(10): a006882.
Biology 2012, 1(3), 521-541; https://doi.org/10.3390/biology1030521

MUNI SCI

HIV reverse transcription

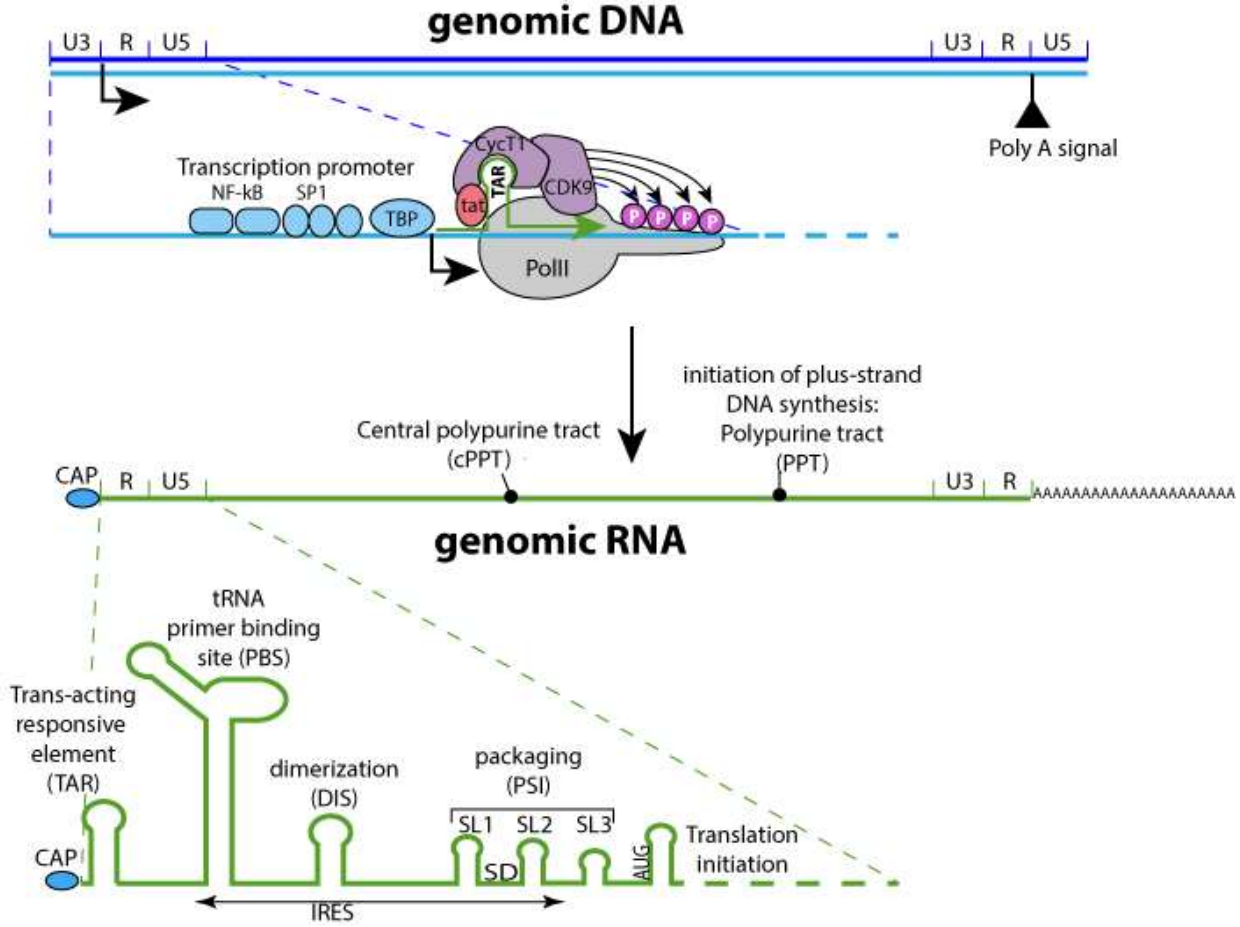
- (E) Plus-strand synthesis continues until the first 18 nucleotides of the tRNA are copied, allowing RNase H cleavage to remove the tRNA primer. Most retroviruses remove the entire tRNA; the RNase H of HIV-1 RT leaves the rA from the 3′ end of the tRNA attached to minus-strand DNA.
- (F) Removal of the tRNA primer sets the stage for the second (plus-strand) transfer.
- (G) Extension of the plus and minus strands leads to the synthesis of the complete double-stranded linear viral DNA.



Cold Spring Harb Perspect Med. 2012 Oct; 2(10): a006882.
Biology 2012, 1(3), 521-541; https://doi.org/10.3390/biology1030521

MUNI SCI

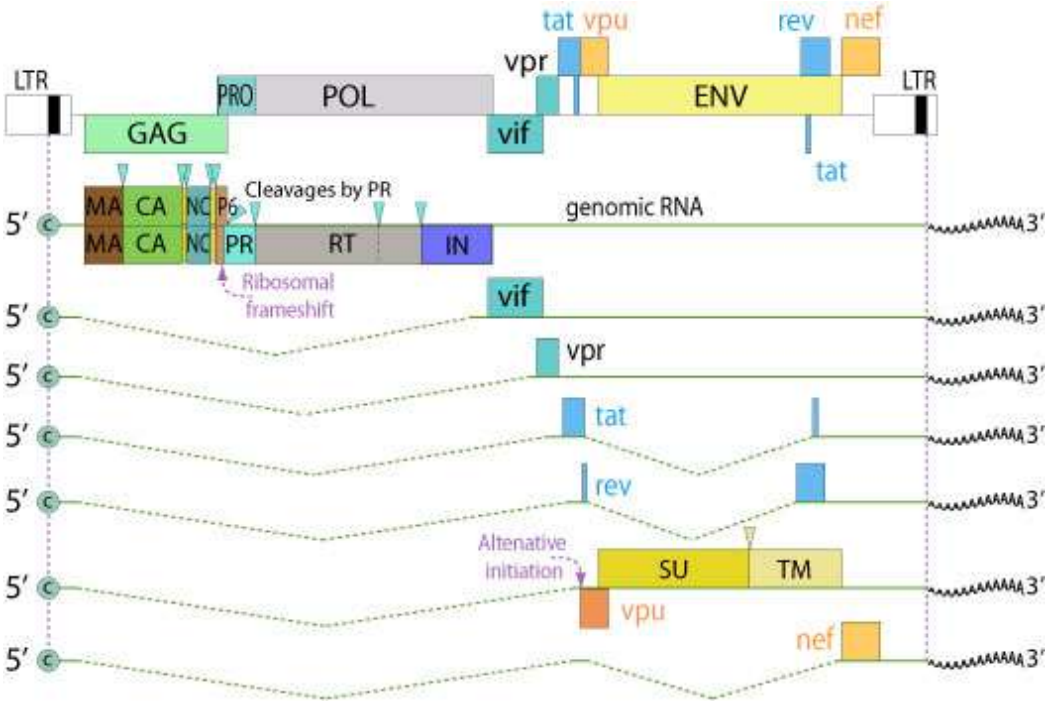
HIV transcription initiation – replication



- Monopartite, linear, dimeric, ssRNA(+) genome of 9,75 kb, with a 5'-cap and a 3'poly-A tail.
- There are two long terminal repeats (LTRs) of about 600 nt long at the 5' and 3' ends.
- The LTRs contain the U3, R, and U5 regions.
- There are also a primer binding site (PBS) at the 5´end and a polypurine tract (PPT) at the 3'end.

HIV transcription - splicing

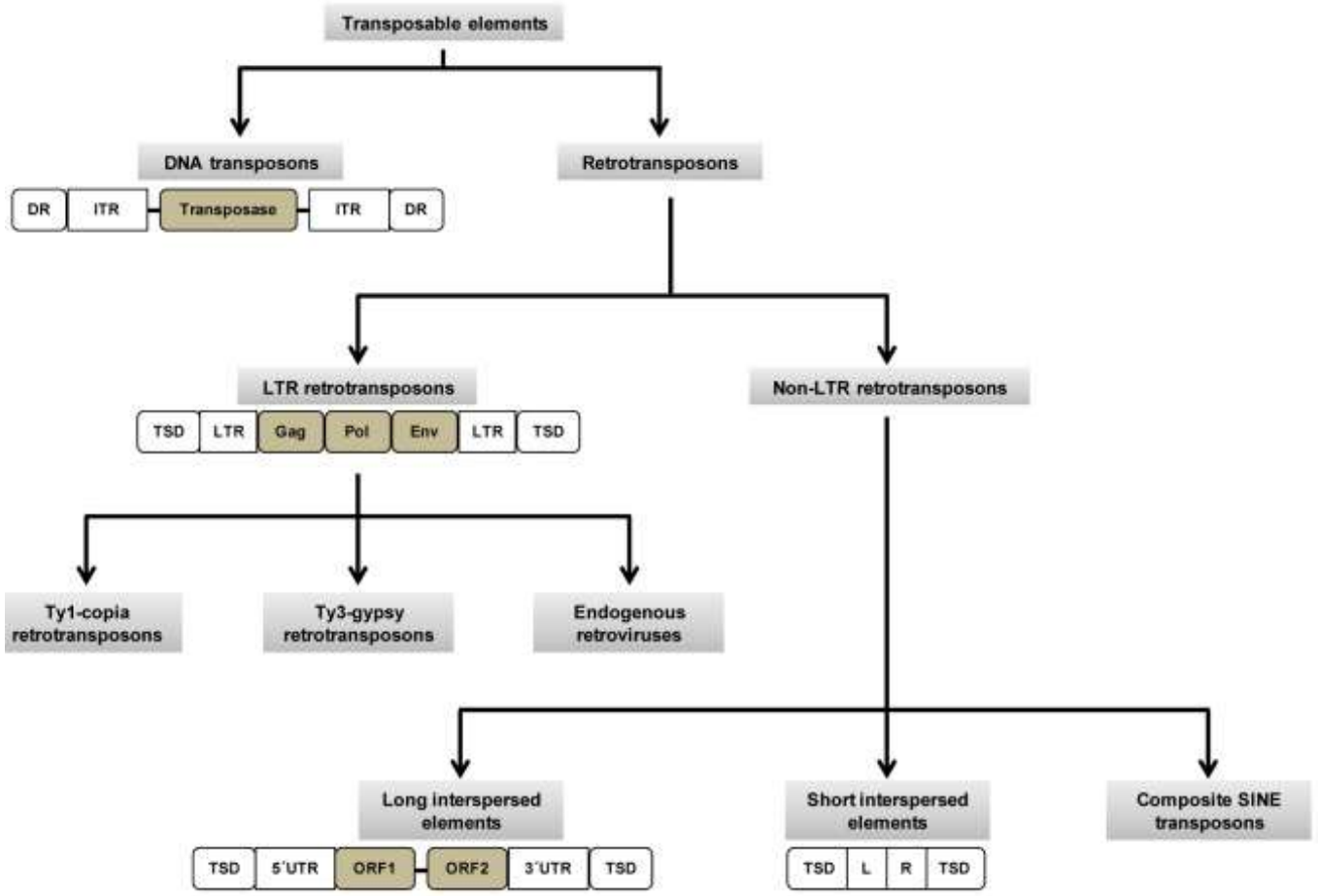
- Unspliced full length mRNA will serve as genomic RNA to be packaged into virions or used as a template for translation of gag and gag-(pro)pol (1 ribosomal frameshift) polyproteins.
- The incompletely spliced mRNAs encode env that is cleaved into SU and TM envelope proteins, and the accessory proteins vif, vpu, and vpr.
- Completely spliced mRNAs encode Rev, Tat and Nef accessory proteins. Rev escorts unspliced and incompletely spliced RNAs out of the nucleus of infected cells.



https://viralzone.expasy.org/5183

MUNI SCI

Transposable elements (TEs)



Genomics Inform. 2014 Sep; 12(3): 98–104.

MUNI SCI

Retroelements – Retrotransposons

- Retrotransposons that replicate through an RNA intermediate and a reverse transcription step, the so-called copy-and-paste transposons, and comprises two main families:

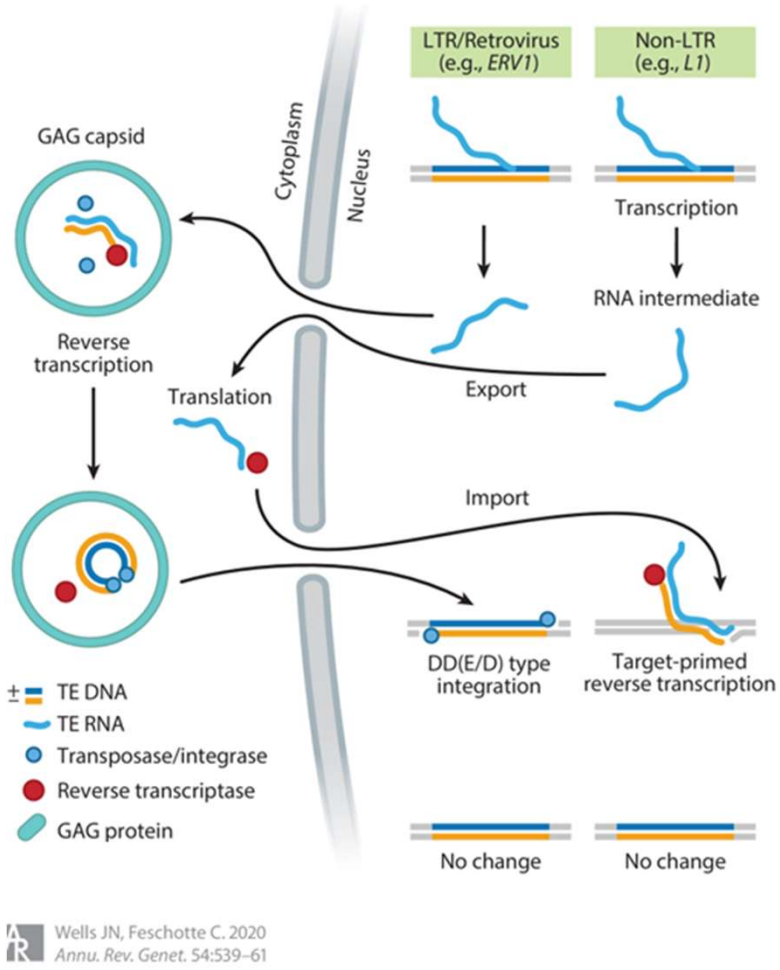
- LTR retrotransposons - contain long terminal repeat - LTR; sometimes specific subclass Endogenous retroviruses (ERVs).

- Non-LTR retrotransposons – without LTR.

- Retrosequences - without LTR, without reverse transcriptase and integrase. Reverse transcripts of structural genes – edited transcripts without introns, with attached poly(A).
  - Retrogenes - functional retrosequences of the original gene coding identical protein.
  - Retropseudogenes - non-functional forms of genes /eg. Alu-sequence in humans (7SL RNA, 300 pb, in humans 500,000 times copied).

LTR and Non-LTR Retrotransposons

- LTR retrotransposons occurs in cytoplasmic virus-like particles and leads to the formation of extrachromosomal double-stranded DNA (dsDNA), which is imported into the nucleus before integrating into a new locus.
- Non-LTR retrotransposons initiate reverse transcription directly at the target locus after cleaving genomic DNA, a process known as 'target-primed reverse transcription'.



Annu. Rev. Genet. 2020. 54:539–61.

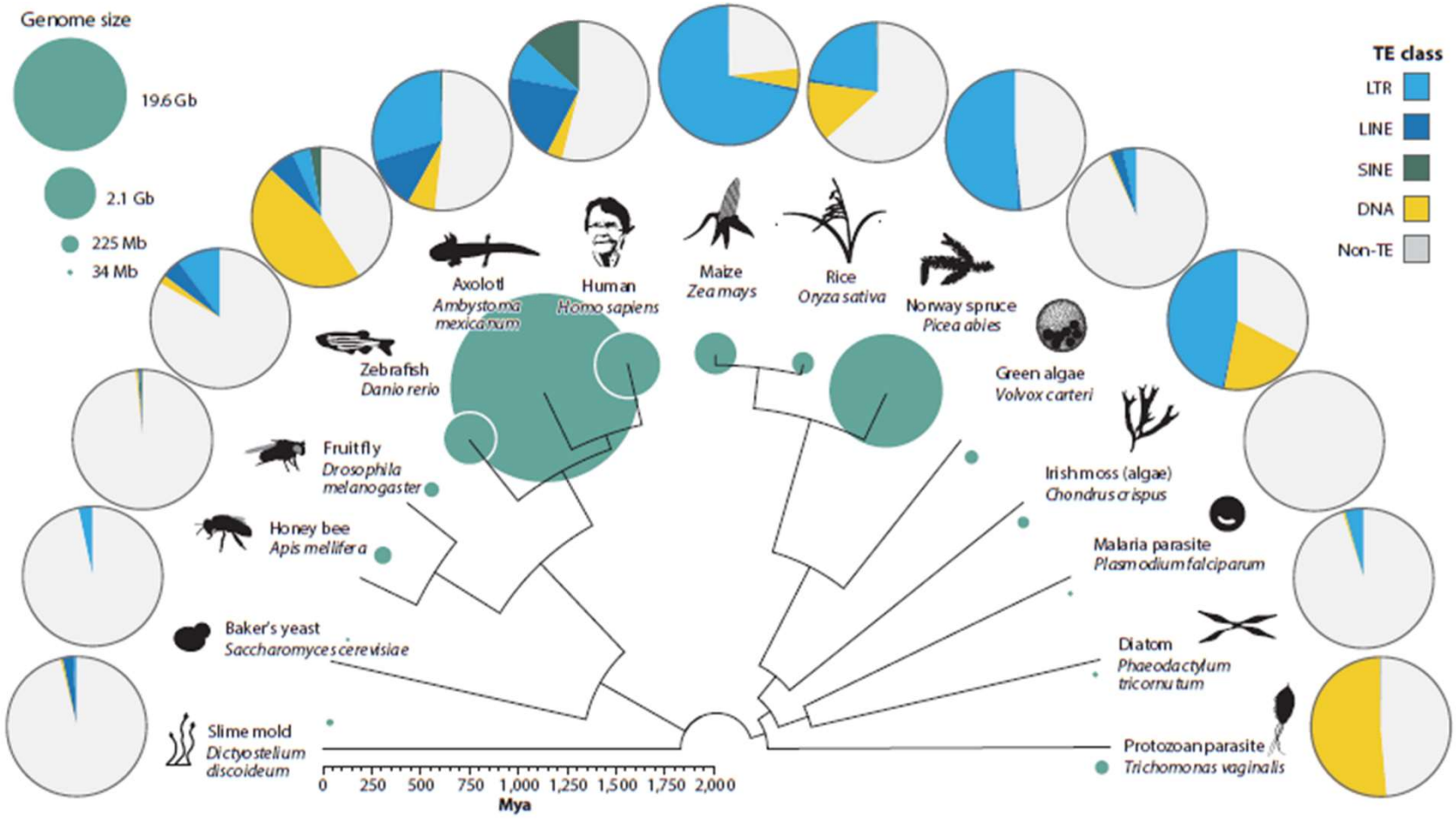
MUNI SCI

Retrotransposons

- LTR retrotransposons occurs in cytoplasmic virus-like particles and leads to the formation of extrachromosomal double-stranded DNA (dsDNA), which is imported into the nucleus before integrating into a new locus.
  - The coding region usually contains only two genes: gag and pol, they do not have env, i.e. virulence factor.
  - Ty-elements in yeast (6.3 kb).
  - Copia-elements a Gypsy-elements in *Drosophila* (5 kb)

- Non-LTR retrotransposons initiate reverse transcription directly at the target locus after cleaving genomic DNA, a process known as 'target-primed reverse transcription'.
  - F, Ga I-elements in *Drosophila*.
  - Short sequences SINE (short interspersed nuclear elements) - 500 bp, 10^5copies, derived from genes for small RNAs, including tRNA (pseudogenes).
  - Long sequences LINE (long interspersed nuclear elements) - 6.5 kb, 10 000 - 50 000 copies in mammals.

Distribution of transposable elements in eukaryotes



Annu. Rev. Genet. 2020. 54:539–61.

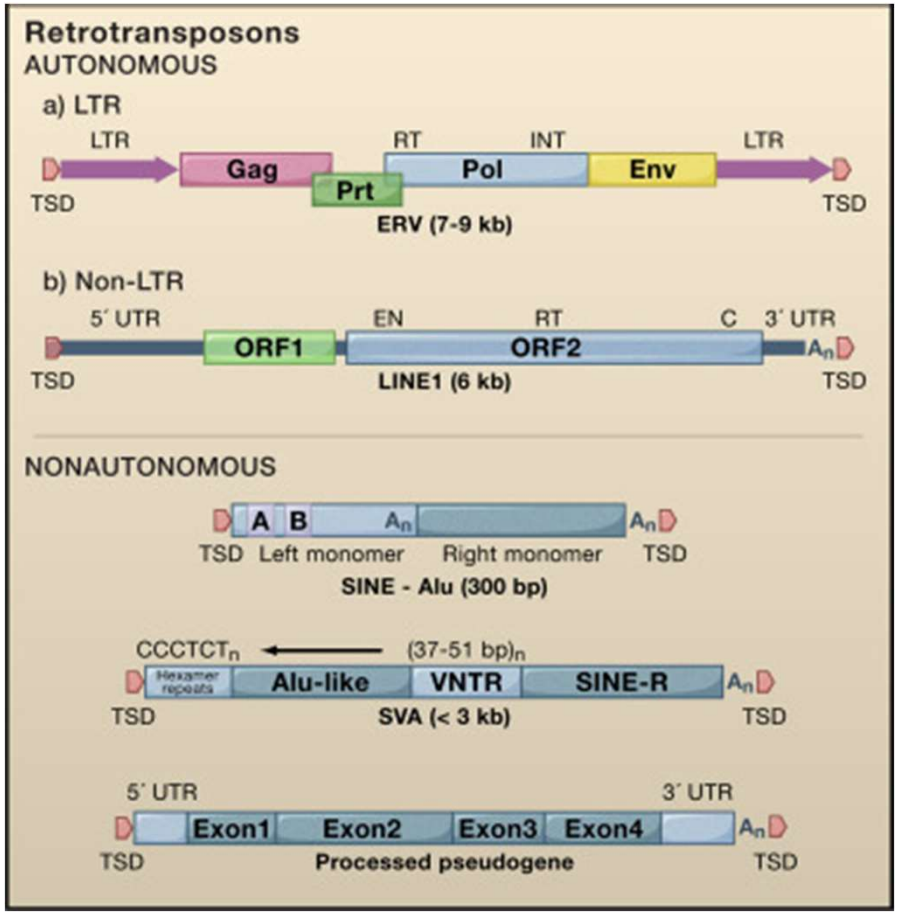
MUNI SCI

Retrotransposons

- Autonomous ERVs and LINEs. The L1 is the only LINE known to be actively mobile in mammals.

- Nonautonomous – Alu and SVA, are dependent on L1 for their mobility. Processed pseudogenes are spliced mRNAs copied and inserted in the genome by the L1s.

- Gag, group-specific antigen (capsid proteins).
- Pol, polymerase.
- Env, envelope.
- LTR, long terminal repeat.
- Prt, protease
- INT, integrase domain.
- RT, reverse transcriptase domain.
- TSD, target site duplication.
- EN, endonuclease domain.



Int J Evol Biol. 2013;2013:424726.

MUNI SCI

Retrotransposons in Yeast

- The most highly characterized Ty1 element is Ty1-H3, which was isolated following its retrotransposition into plasmid DNA.

- Resembles a primitive retrovirus.
- Approximately 35 copies per haploid yeast genome.

- 5918 base pairs (bp) in length with 334 bp direct repeats, or LTRs, at each end, with 5 bp duplication when incorporated.

- Ty1 – element has two genes – *TyA* and *TyB*, which are homologous to the Gag and Pol retrovirus genes.

- Products of *TyA* and *TyB* form virus-like particles in the cytoplasm.

Retrotransposons in Yeast

- LTRs - boxed arrowheads.
- Functional domains of Pol that synthesized as part of the Gag-Pol polyprotein are posttranslationally cleaved by PR (protease) into separate proteins include.
- RT-RH (reverse transcriptase-RNase H).
- IN (integrase).
- The retroviral envelope gene (ENV) is not present in Ty1.

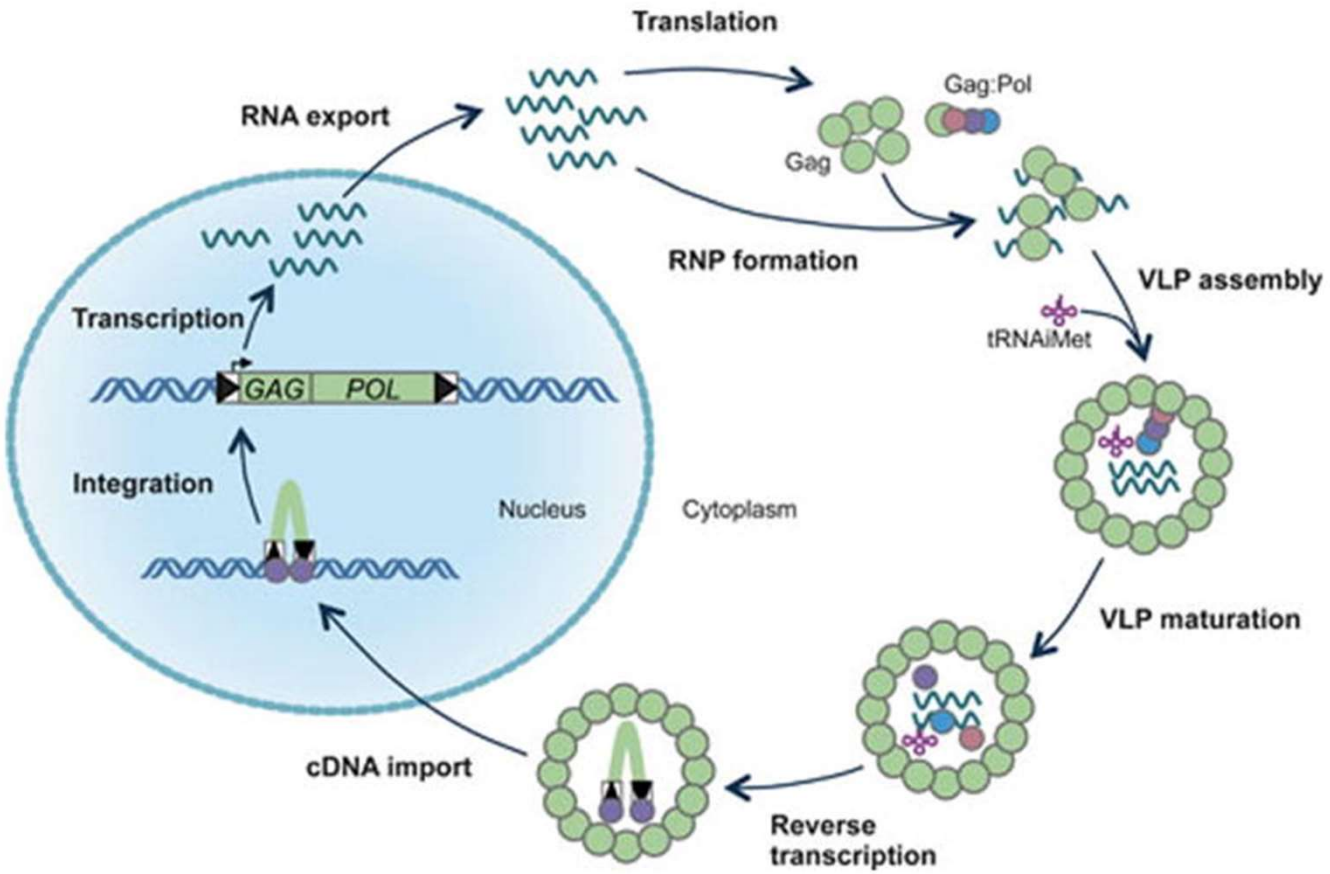
Microbiol Spectr. 2015 Apr 1; 3(2): 1–35.

Mechanism of retrotransposition of Ty element

- A Ty1 element in the host genome is transcribed and the Ty1 RNA is exported to the cytoplasm.
- The RNA is translated into Gag and Gag-Pol proteins, and associates with these proteins to form Ty1 RNPs, also known as retrosomes.
- Ty1 RNPs give rise to VLPs that encapsidate a dimer of Ty1 RNA and $tRNA^{iMet}$.
- Within the VLP, Gag and Pol proteins are cleaved by PR to form mature Gag, PR, IN and RT proteins.
- Following VLP maturation, Ty1 RNA is reverse transcribed into cDNA by RT using $tRNA^{iMet}$ as a primer. The cDNA is bound by IN to form the pre-integation complex, which is imported into the nucleus.
- IN integrates Ty1 cDNA into the yeast genome.
- At the insertion site of DNA there is duplication of 5 nucleotides similar to bacterial transposons.

Mechanism of retrotransposition of Ty element

- Transcription and Ty1 RNA is exported to the cytoplasm.
- The RNA is translated into Gag and Gag-Pol.
- Formation of Ty1 RNPs retrosomes.
- VLP assembly - encapsidation a dimer of Ty1 RNA and $tRNA^{iMet}$.
- Cleavage of Gag-Pol into Gag, PR, IN and RT.
- Following VLP maturation - reverse transcription using $tRNA^{iMet}$ as a primer.
- IN integrates Ty1 cDNA into the yeast genome.



Microbiol Spectr. 2015 Apr 1; 3(2): 1–35.

MUNI SCI

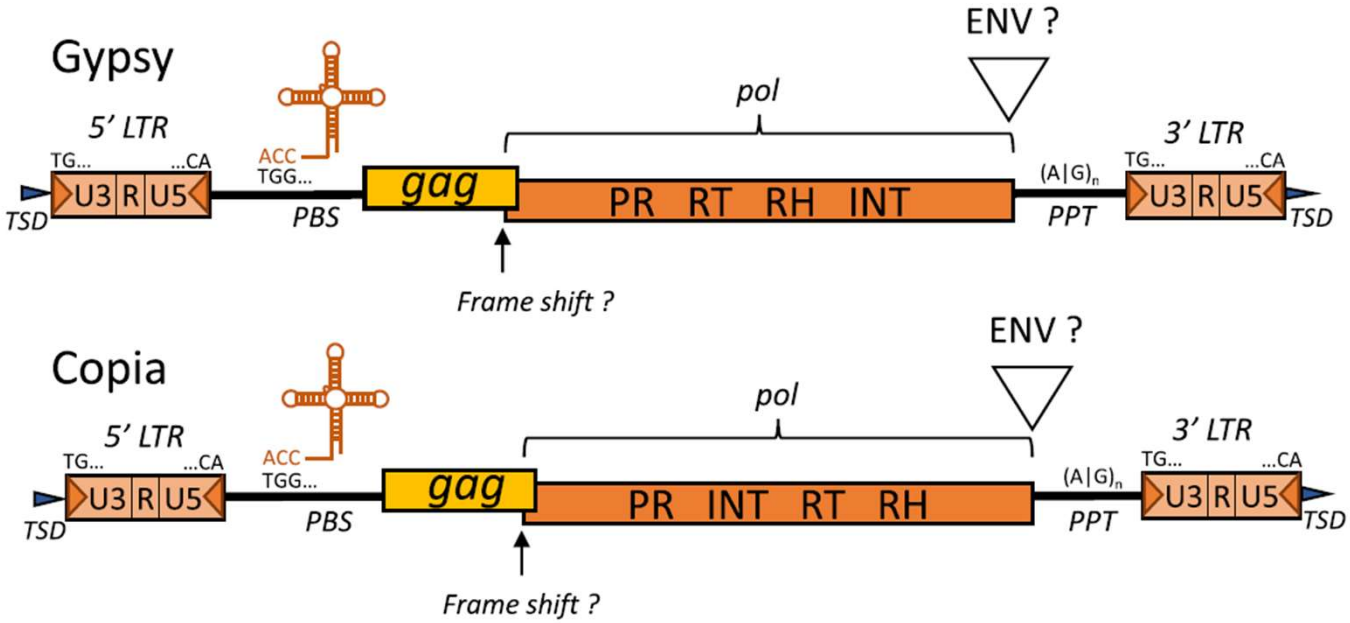
Retrotransposons in *Drosophila melanogaster*

- LTR retrotransposons:
  - Copia elements – produce a large amount of RNA (hence the name), structurally similar to Ty1-yeast elements.
  - Gypsy elements – larger than Copia, they also contain a gene similar to the *Env* gene in retroviruses.
  - Copia and Gypsy forms virus-like particles in drosophila cells; however, only particles containing gypsy RNA can pass through the cell membrane.

- Non-LTR transposons:
  - The F, G and I-elements - they do not have LTR, at the ends they have sequences formed by reverse transcription of poly(A).
  - HeT-A and TART (telomere-associated retrotransposon).

Oncotarget, February 2019 10(12).

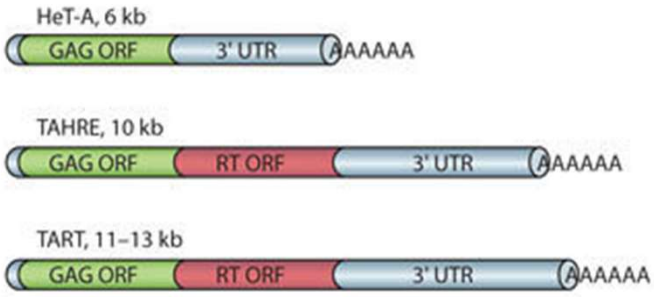
Retrotransposons in *Drosophila melanogaster*



MUNI SCI

Het-A and TART extend telomeres in *Drosophila*

Non-LTR elemetns



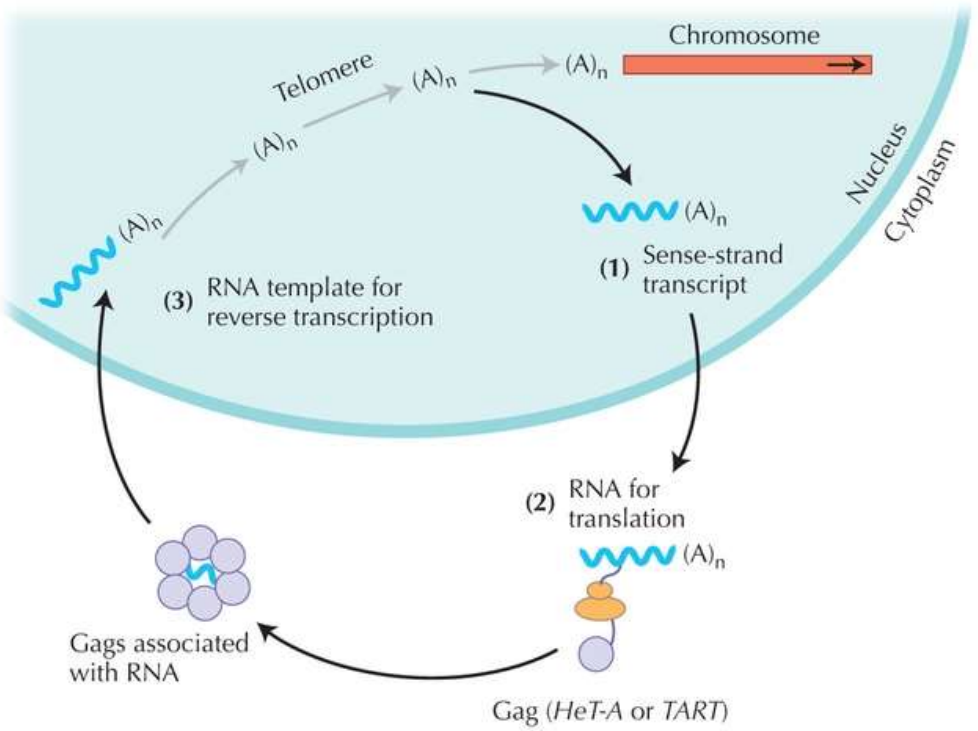
Telomere structure

- Prolong telomere sequences.
- Specifically incorporate into the positions at the ends of chromosomes.
- These elements carry long 3′ UTRs of at least 3 kb. The 3′ oligo(A) tail used to attach to chromosome ends is indicated by AAAAAA.
- Drosophila telomers thus include a tandem mixed array of variably 5′ truncated retrotransposons.
- The ‘A’ at each junction indicates the 3′ oligo(A) tail.
- Termed telomere associated sequence (TAS) is followed by unique sequence chromosomal DNA.

Cytogenetic and Genome Research, Fabruary 2008, 122(3-4):356-64

Het-A and TART extend telomeres in *Drosophila*

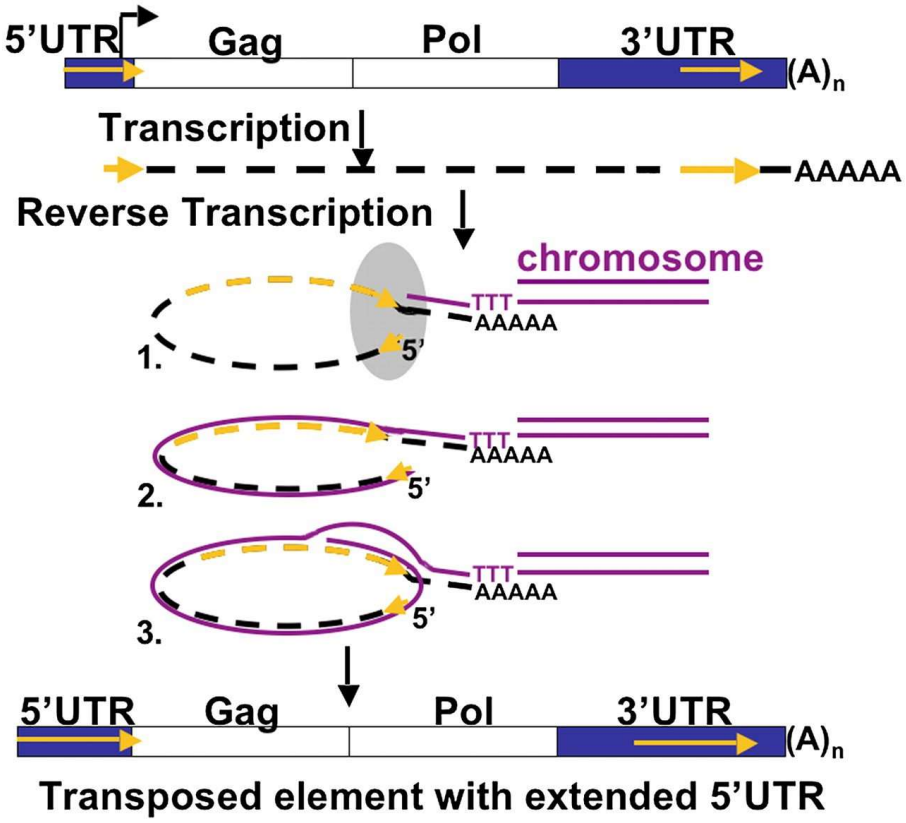
- Drosophila telomeres consist of long arrays of two non-LTR retrotransposons, HeT-A and TART. Addition of these elements maintains the telomere despite its tendency to shorten during replication.
- These retrotransposons produce a sense-strand transcript with a poly(A) tail, denoted (A)n.
- This is transported to the cytoplasm, translated to produce Gag proteins that remain associated with the RNA.
- Gag associated with RNA is transported back to the nucleus, and reverse transcribed to add an extra copy specifically to the end of the telomere.



http://www.evolution-textbook.org/content/free/notes/ch21_WN21D.html

MUNI SCI

Het-A and TART extend telomeres in Drosophila

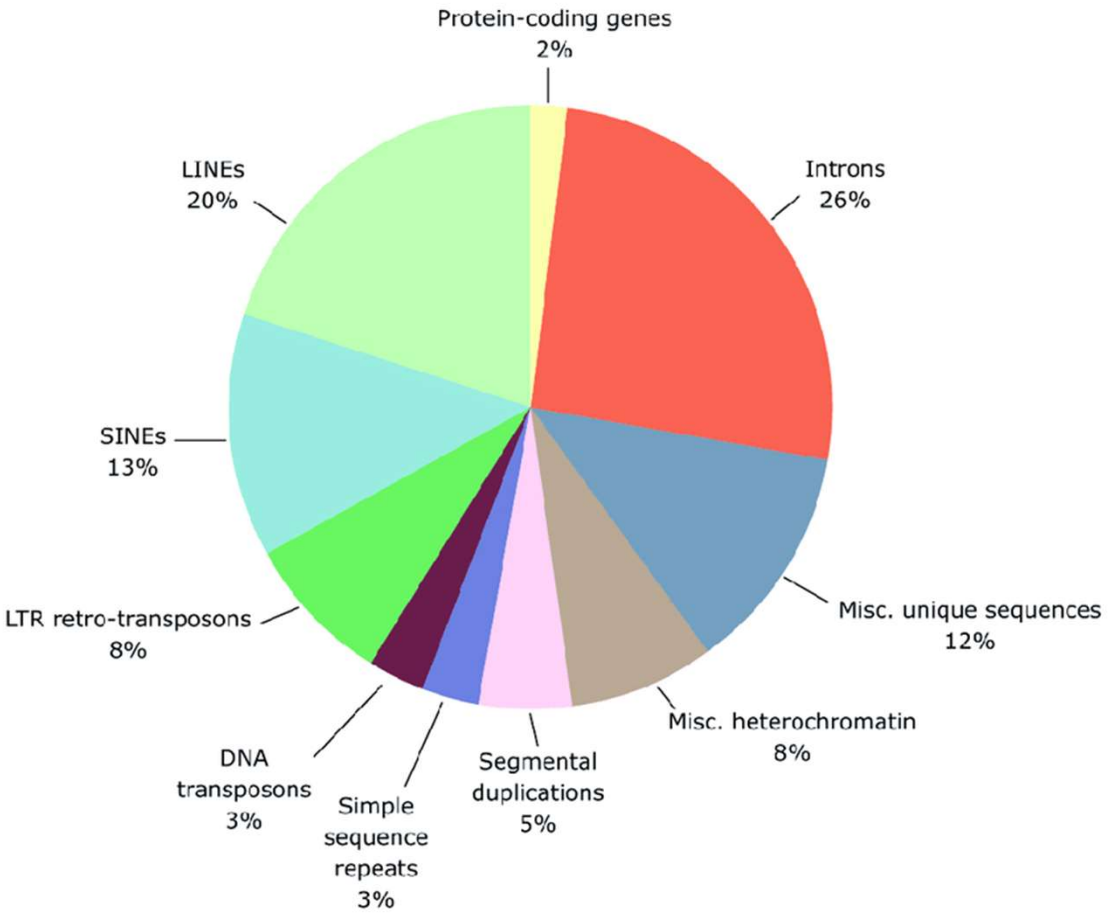


- 1. The polyA tail associates with the chromosomal DNA, and RT begins to copy the RNA. Protein complex brings the 5′ PNTR sequence into proximity to the 3′ end of the 3′ PNTR.
- 2. When RT reaches the 5′ end of the transcript, it makes a template jump back to the matching 3′ end of the 3′ PNTR.
- 3. RT dissociates the RNA–DNA complex and recopies some or all of the 3′ PNTR.
- As a result, the transposed element will have more 5′ UTR sequence than the RNA.

PNAS, August, 2011, 108 (51) 20317-20324.

Composition of human genome

- Protein-coding genes make up about 2% of human genomic DNA.
- Different types of repetitive DNAs make up more than 50% of the genome.
- These include, in particular, mobile elements.



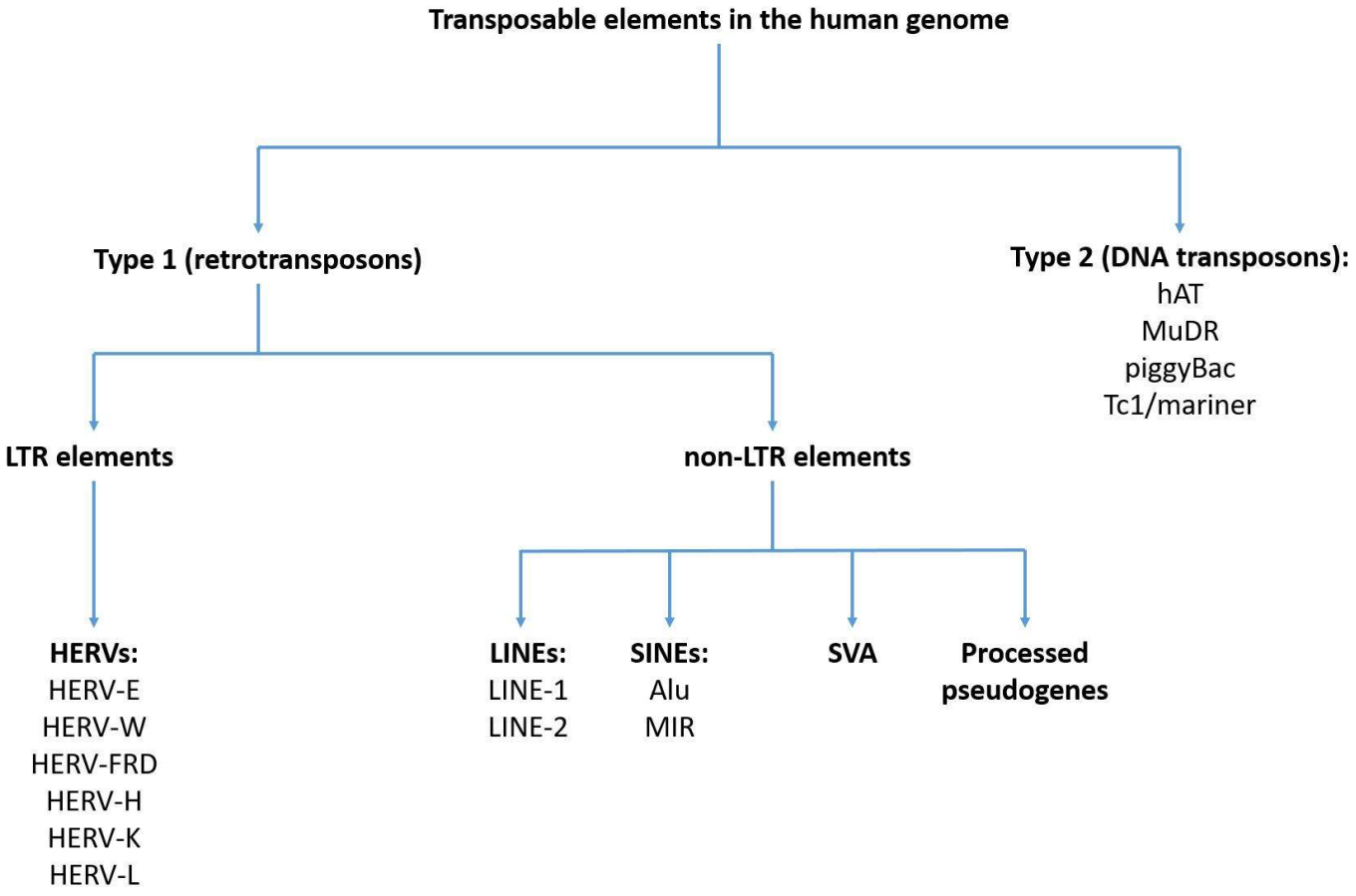
MUNI
SCI

Repetitive sequences in human genome

- VNTR - variable number of tandem repeats - copies following one after the other in a certain locus of the genome.
  - Macrosatellite (heterochromatin regions in the centromere region).
  - Minisatellite (repetition of the sequence 5 –30 bp).
  - Microsatellite (identification of persons).
  - Telomeric (maintained telomeres at the ends of chromosomes).

- Dispersed - scattered throughout the genome, mostly capable of transposition:
  - LINE
  - SINE
  - LTR retrotransposons - replicative mechanism of transposition.
  - DNA transposons - mechanism „cut and paste“.

Classification of TE in human genome

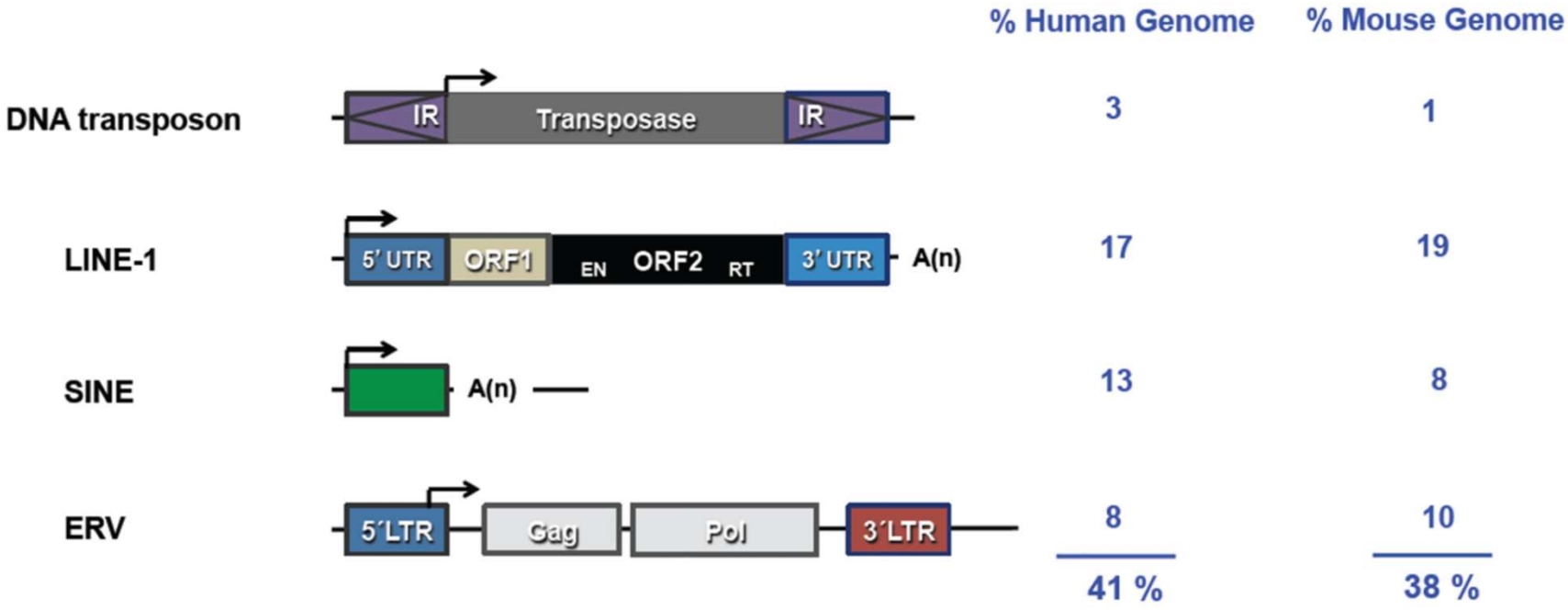


Misiak B, et al. Front Genet. 2019. PMID: 31293617

MUNI SCI

Repetitive sequences in human genome

- At least 44% of human DNA is derived from transposable elements:
- Engndogenous retrovioruses - 8%.
- Non-LTR retrotransposon – 30 %.
- DNA transposons - cut and paste' elements – 3 %.



Development. 2016 November 15; 143(22): 4101–4114.

MUNI SCI

SINE elements

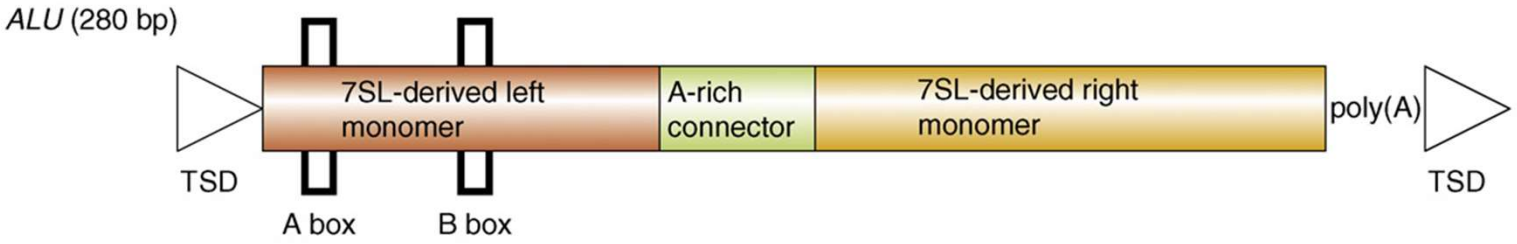
- Short interspersed nuclear elements.
- Non-LTR retrotransposons and Nonautonomous.
- Size 100 -700 pb (other sources: less than 400 bp).
- The internal sequences of SINE are conservative, derived from tRNA-encoding genes (they do not code any polypeptide).
- Widespread in eukaryotes (13% of the human genome). Structurally species-specific, macroervolutionary divergence.
- Common development with LINE, because SINE do not encode reverse transcriptase and without LINE they cannot move (parasite of parasites??).
- Transcribed by RNA-polymerase III from its own promoter.
- Significant contribution to genome plasticity, regulation of gene expression.

- In the human genome of 3 SINE families:
  - Alu (only Alu capable of transposition).
  - MIR.
  - Ther/MIR3.

Int J Evol Biol., 2013;2013:424726.

Alu repeat or sequence

- The most common type of SINE.
- Size 28 - 350 bp.
- Contain a target site for restriction enzyme Alu I.
- Make up 10-11% of the human genome (1-1.5 million copies).
- Long thought to be junk" DNA with no function, but apparently involved in some cellular processes:
  - Forms the place of attachment of cohesin complexes of replicated chromosomes before segregation.
  - Many individual Alu elements have wide-ranging influences on gene expression - polyadenylation, splicing and ADAR (adenosine deaminase that acts on RNA) editing.

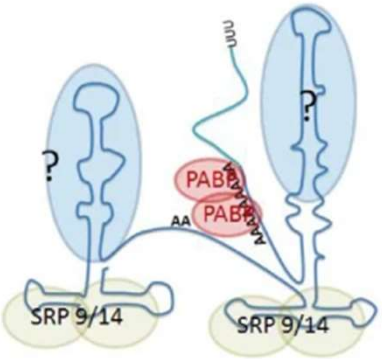


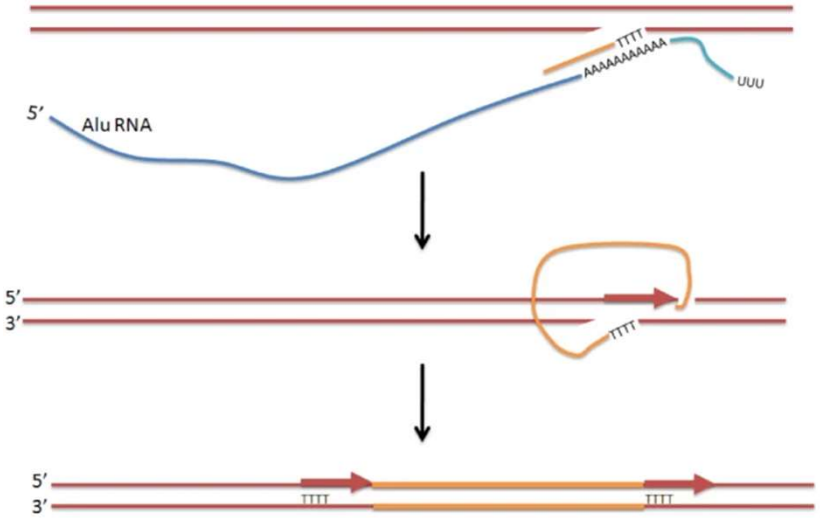
- Boxes A and B are internal promoters for RNA polymerase III.

Review Trends Genet., 2007 Apr;23(4):183-91.

Alu repeat or sequence

- The Alu RNA has been shown to bind the 7SL RNA SRP9 and 14 heterodimer, as well as polyA-binding protein (PABP).



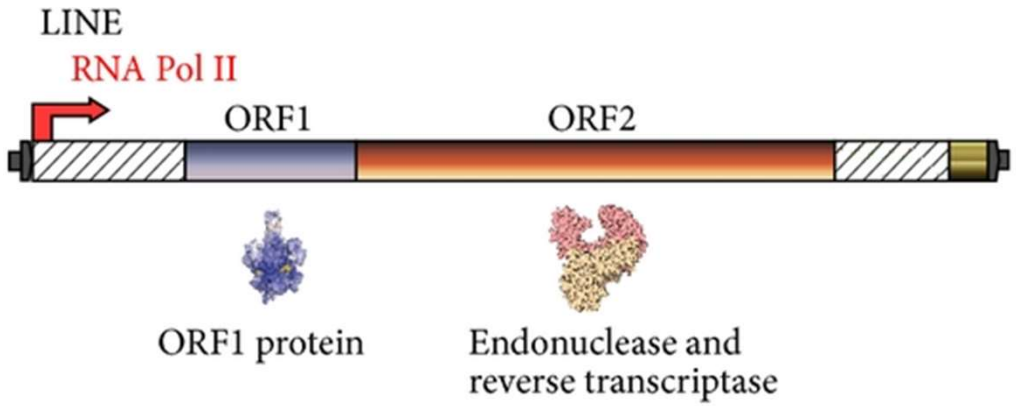


- The Alu RNA brings the ORF2p to the genome where its endonuclease activity cleaves at a T-rich consensus sequence.
- The T-rich region primes reverse transcription by ORF2p on the 3' A-tail region of the Alu element.
- A nick occurs by an unknown mechanism on the second strand and second-strand synthesis is primed.
- The new Alu element is then flanked by short direct repeats that are duplicates of the DNA sequence between the first and second nicks.

Genome Biology, 2011, Volume 12, Article number: 236.

LINE elements

- Long interspersed nuclear elements.
- Non-LTR retrotransposons and Autonomous (express proteins for its own transposition).
- Size approx. 7000 pb.
- Transcribed by RNA Pol II from its own promoter.
- Make up around 20% of the human genome.
- Encode at least one ORF2 protein: reverse transcriptase with endonuclease domain to ensure the formation of a DNA copy of LINE and its incorporation into the genome.
- In the human genome, there are about 850,000 copies of complete LINE elements and truncated mutant copies, that are no longer subject to transcription/translation.



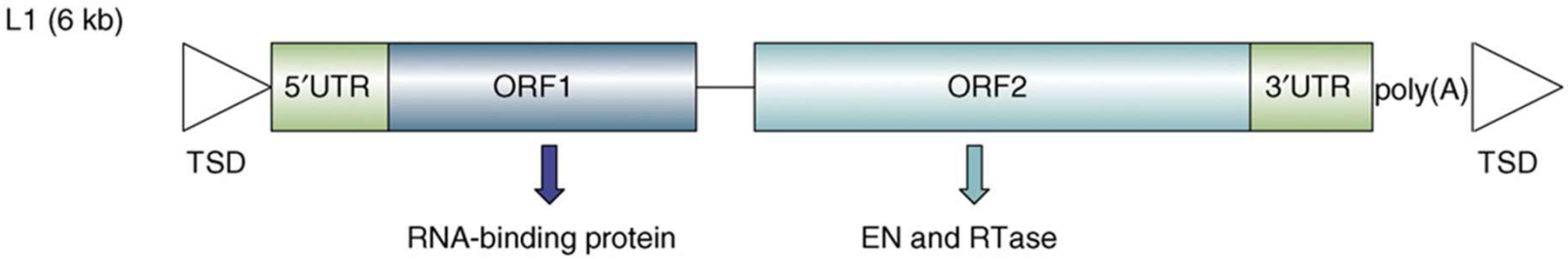
Int J Evol Biol., 2013;2013:424726.

Retrotransposon L1-element

- About 6 kb long.
- The 5' UTR of the L1 element contains a strong, internal RNA Polymerase II transcription promoter in sense.

- ORF1 - encodes RNA and RNA-binding protein and ORF2 - encodes a protein with reverse transcriptase and endonuclease activity.

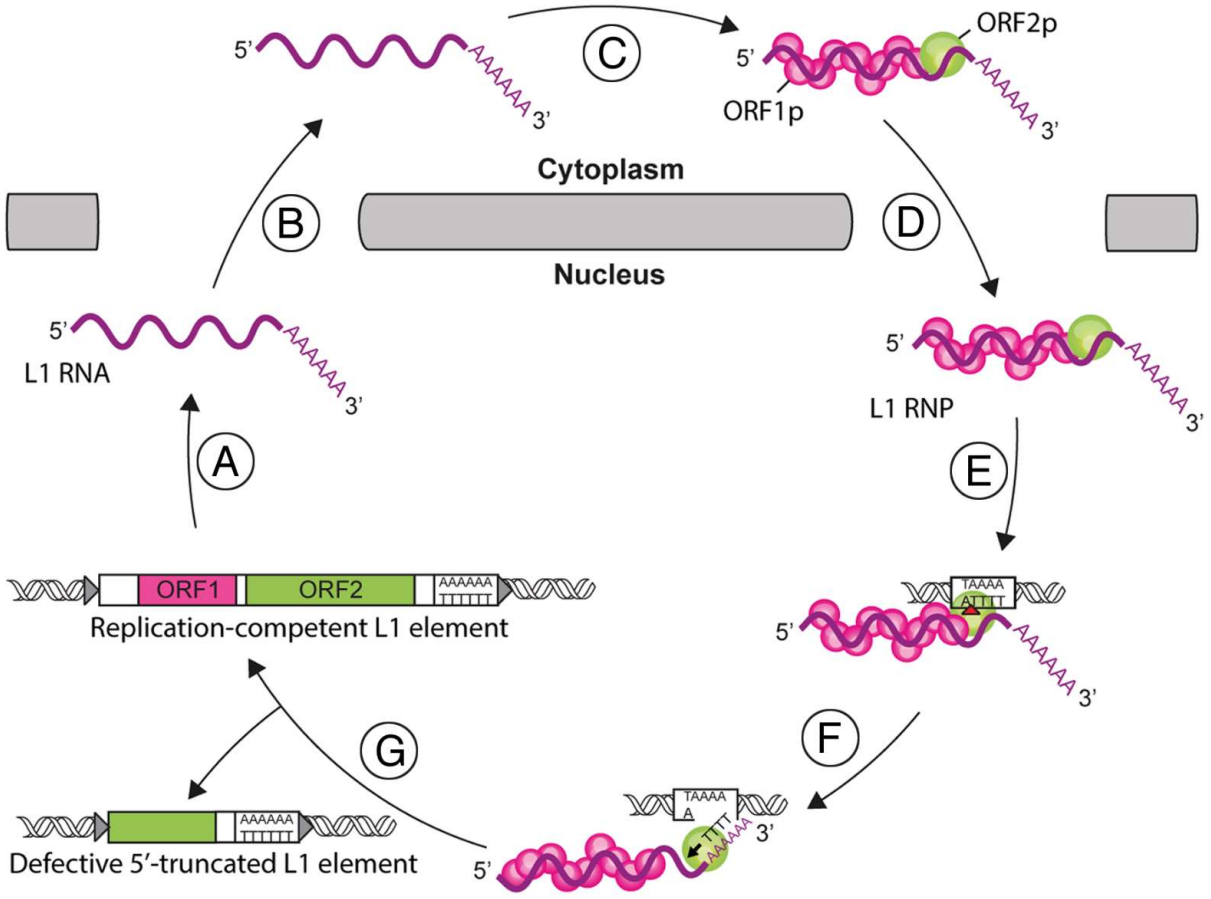
- The human genome contains about 3000 to 5000 complete L1-elements and about 500,000 L1-elements truncated to 5'-rings that do not have the ability to transpose; all elements are bounded by a short duplication (TSD) of the destination.



Review Trends Genet., 2007 Apr;23(4):183-91.

Replication of L1-element

- (A) L1 replication starts by the transcription of a bicistronic mRNA.
- (B) The L1 RNA is exported to the cytoplasm.
- (C) ORF1p, RNA-binding protein, and ORF2p, endonuclease and reverse transcriptase, are translated and bind to the L1 RNA to form L1 ribonucleoprotein particles (RNP).
- (D) The L1 RNP is imported into the nucleus.

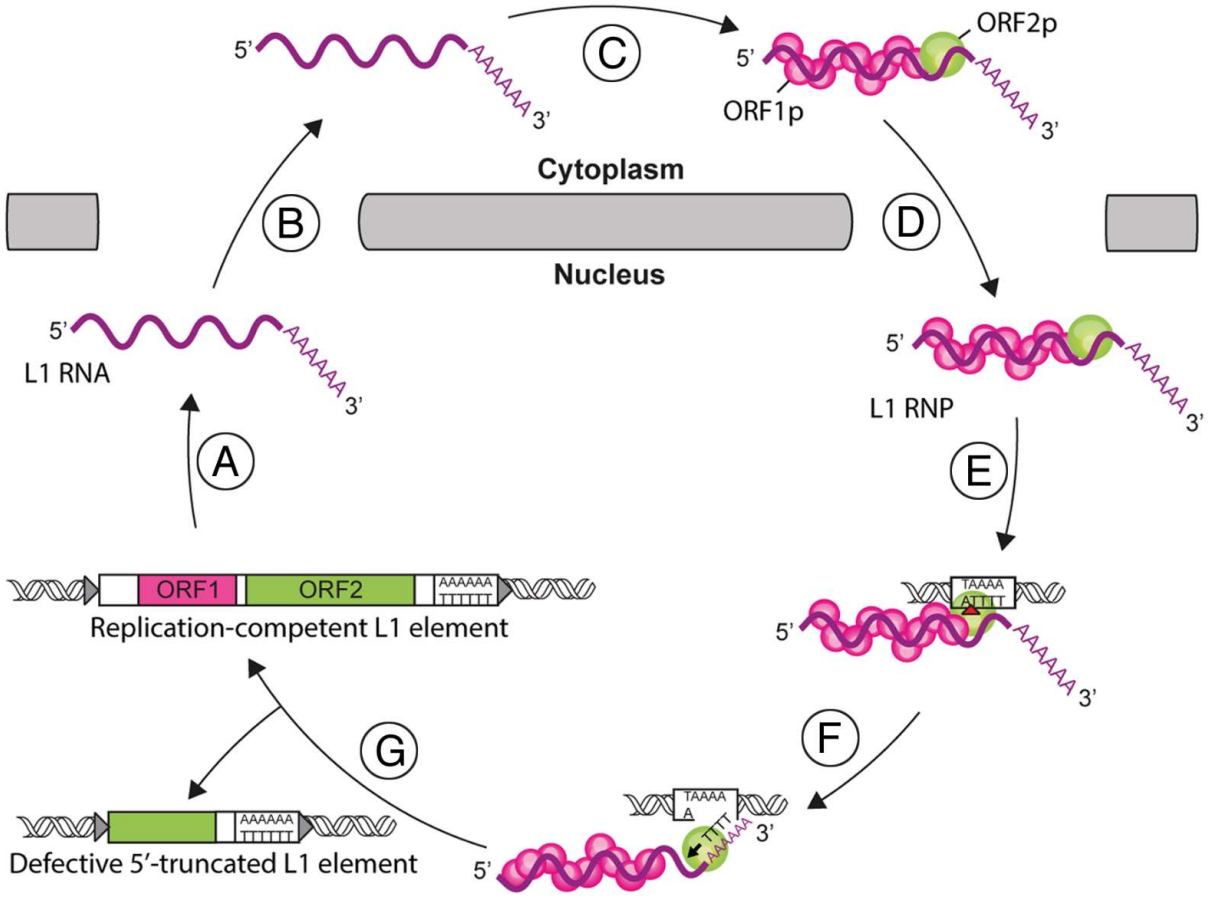


http://eul1db.ircan.org/references.jsp

MUNI SCI

Replication of L1-element

- (E) First, the L1 endonuclease (EN) activity nicks the target DNA (red arrowhead, E). Integration.
- (F) Then, the L1 reverse transcriptase (RT) initiates the reverse transcription of L1 RNA through annealing between the target site and the poly(A) tail of the L1 RNA (black arrowhead, F).
- (G) The mechanisms involved in the final steps of this process and the resolution of the integration are unresolved yet. Partial reverse transcription can lead to 5'-truncated L1 copies.



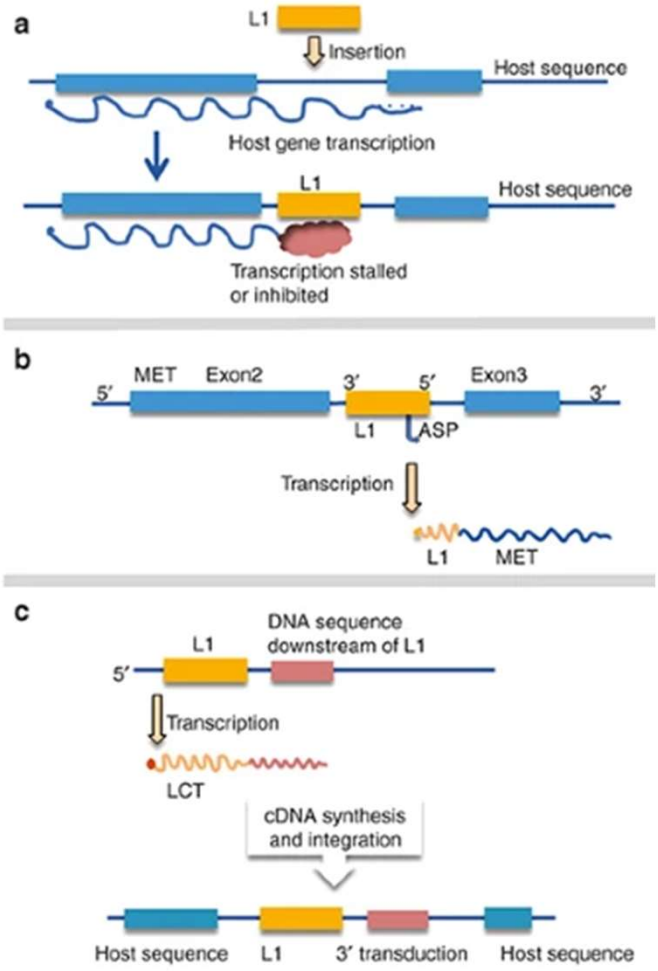
http://eul1db.ircan.org/references.jsp

MUNI SCI

Impact of L1-element transposition

- Transposed copies of complete L1 elements were detected in the analysis of individuals with genetic diseases such as hemophilia and muscular dystrophy.
- Fortunately, these aberrations are rare, indicating that the frequency of L1 transposition is low.
- There are other types of LINE sequences in the human genome:
  - 315,000 copies of L2 (not able to transpose).
  - 37,000 copies of L3 (not able to transpose).

Impact of L1-element transposition



- (a) L1 insertion-mediated inhibition of host gene transcription: L1 can potentially act to slow RNA pol II elongation, dissociate it from the template, or induce premature termination of transcription.
- (b) L1 insertion–mediated oncogene activation: the ASP within L1 inserted antisense to gene MET serves as a transcription start site to drive MET expression.
- (c) 3′ transduction: downstream sequence of L1 3′ end is transcribed together with L1 and the resultant LCT is reverse-transcribed and integrated into a new locus by L1 retrotransposition machinery.

Genetics in Medicine, 2016, volume 18, pages431–439 (2016)

MUNI SCI

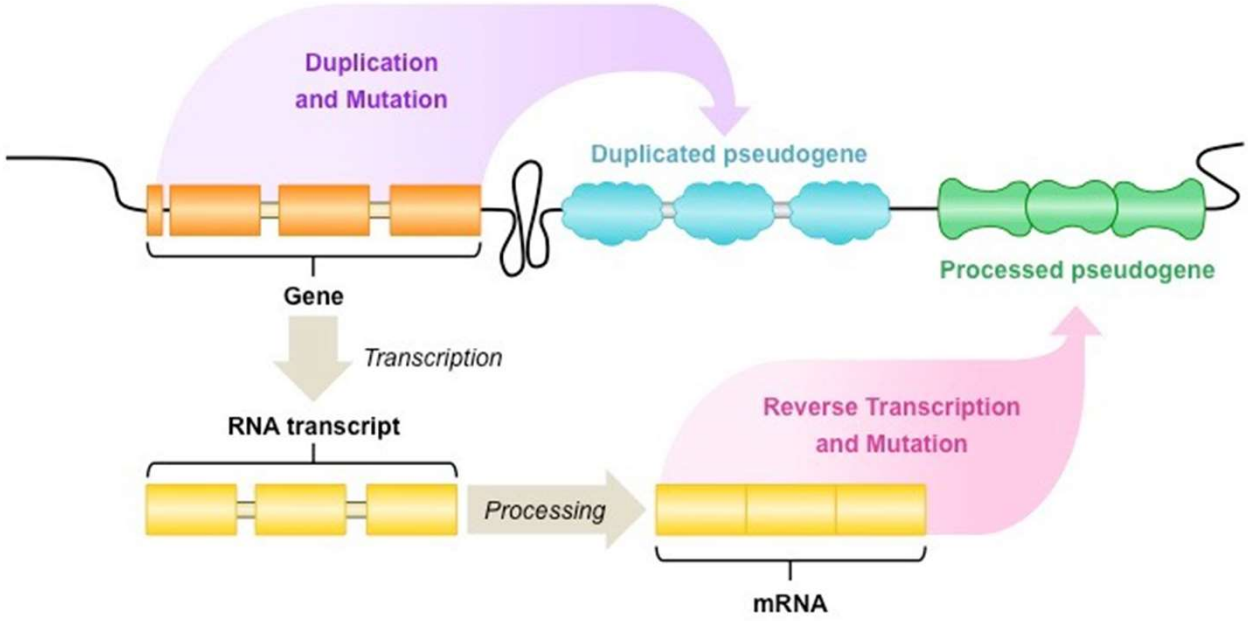
Pseudogenes

- A pseudogene is a segment of DNA that structurally resembles a gene but is not capable of coding for a protein.
- Pseudogenes are most often derived from genes that have lost their protein-coding ability due to accumulated mutations that have occurred over the course of evolution.
- Consequence of mutation(s) of the parental gene.
- If another gene takes over the lost function, the pseudogenes in the genome are maintained and accumulate additional mutations.
- Suitable material for studying random changes in DNA sequences over time.

https://www.genome.gov/genetics-glossary/Pseudogene

Pseudogenes

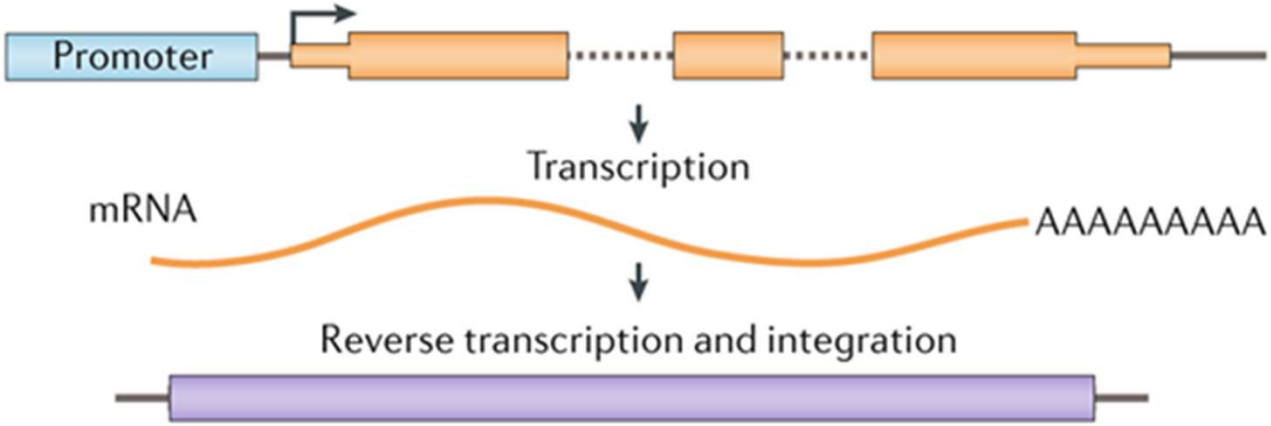
- Most pseudogenes arise from the duplication - carry introns.
- Modified (processed) pseudogenes are generated by transcription of mRNA into the genome by integration - do not contain introns.



https://ib.bioninja.com.au/standard-level/topic-3-genetics/31-genes/pseudogenes.html

Processed pseudogenes

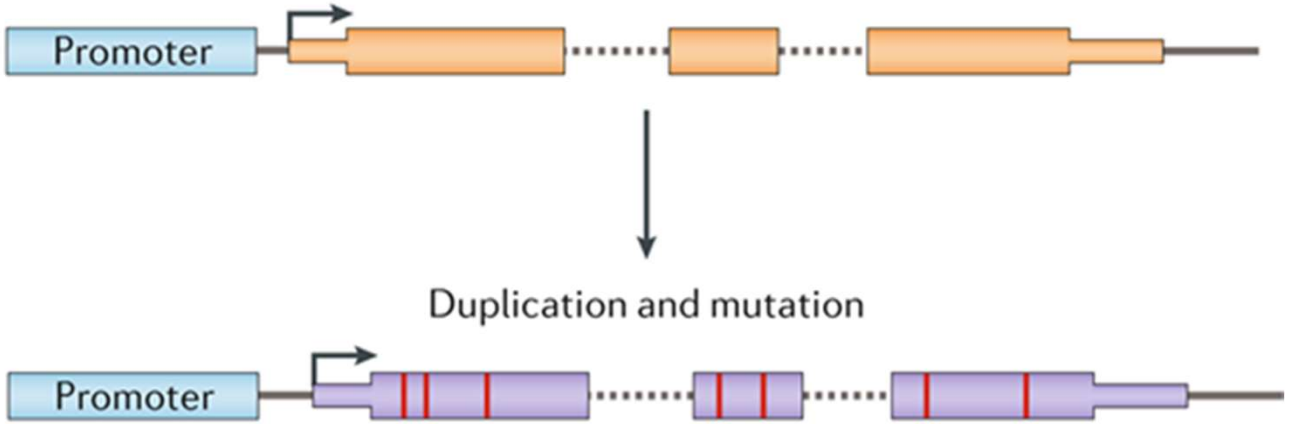
- Processed pseudogenes arise from the reverse transcription and integration of a processed mRNA. Processed pseudogenes consequently lack introns and promoter sequences, but may include a poly-A tract. These pseudogenes may be randomly integrated anywhere in the genome.



Nature Reviews Genetics, 2020, Volume 21, pages191–201.

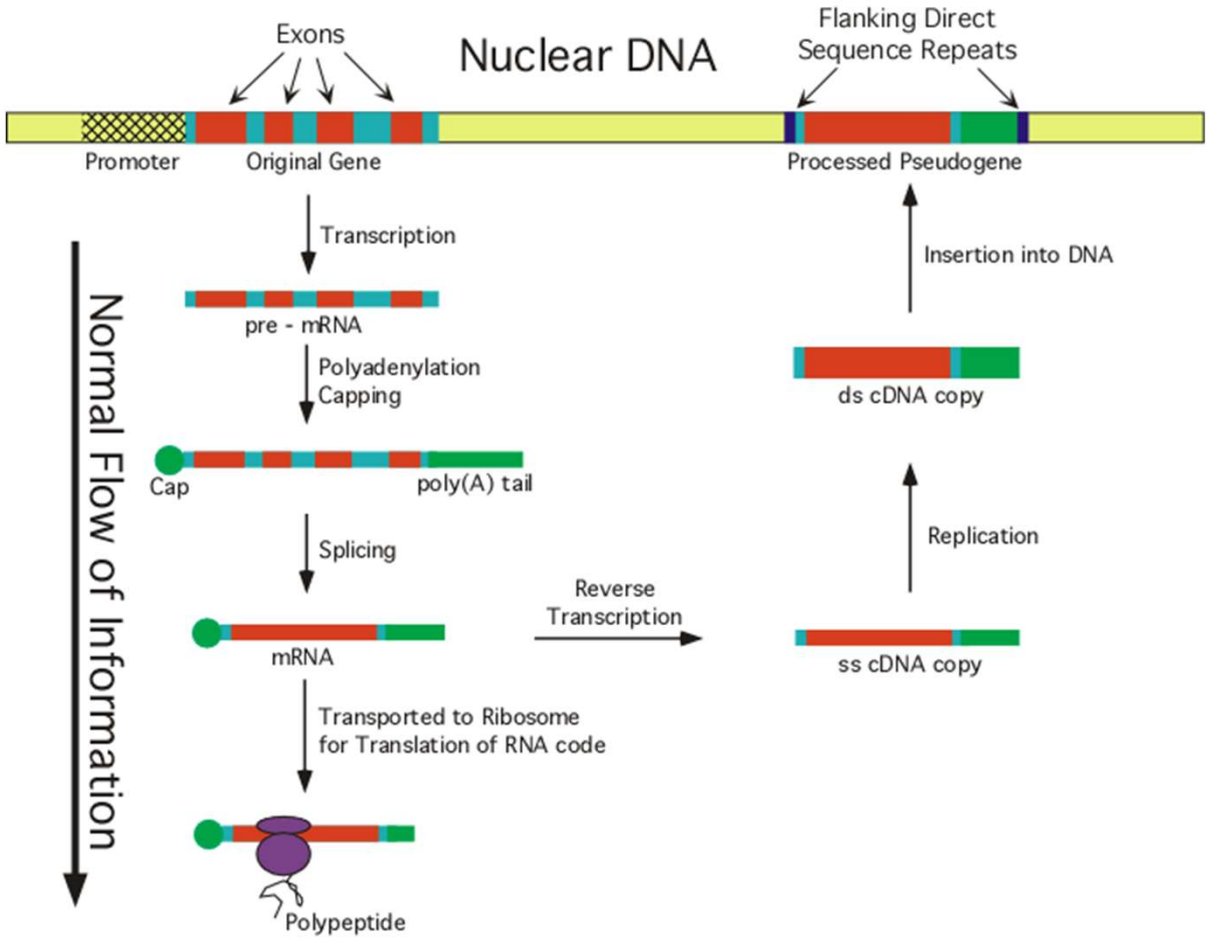
Unprocessed pseudogenes

- Unprocessed pseudogenes originate from gene duplications that accumulate mutations, preventing their translation. Non-processed pseudogenes will often be flanked by transcriptional regulatory elements (e.g. promoters, etc.) These pseudogenes are usually located adjacent to the original gene.



Nature Reviews Genetics, 2020, Volume 21, pages191–201.

Establishment of new pseudogenes

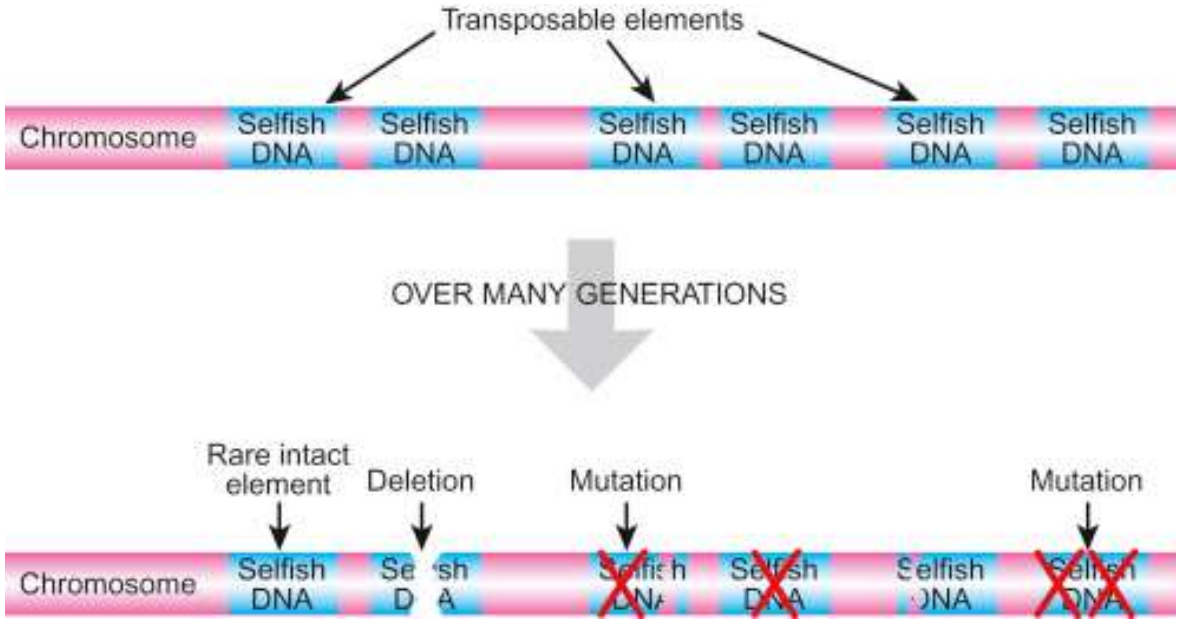


MUNI
SCI

Selfish DNA

- Transposable elements have been called "junk" DNA and "selfish" DNA.
  - "selfish" because their only function seems to make more copies of themselves.
  - "junk" because there is no obvious benefit to their host.

- Over time, many copies of selfish DNA are inactivated by mutations and deletions, leaving DNA remnants called junk DNA.



http://grupo.us.es/gfnl/dna/genetic_ingeniering/transposons.htm

https://www.sciencedirect.com/topics/neuroscience/selfish-dna

MUNI SCI

Jumping genes – helpers or destroyers?

- Transposons found so far in all studied genomes.
- Genomic parasites - Francis Crick, 80s of the 20th century.
- The reverse transcriptase gene is the most numerous gene in the human genome.
- The largest family of human retrotransposons Alu (1.5 million copies): reintegration occurs in every twentieth newly born individual.
- During brain development, there are thousands of "jumps" of retrotransposons L1 and Alu to new locations - the brain is a mosaic of genetically different neurons!
- Transposon activity is strongly increased in tumor cells.

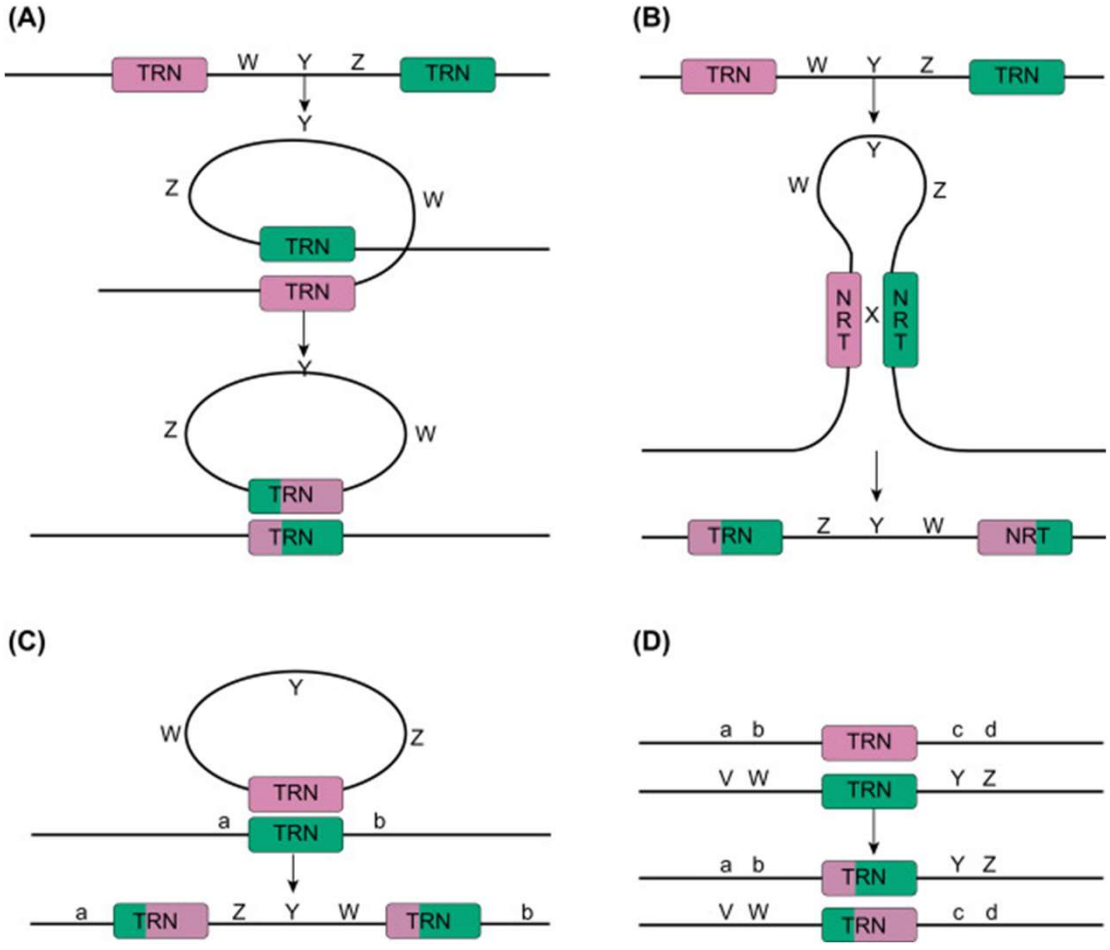
Function of transposable elements

- Provide a selection advantage to their hosts.
- Act as genome parasites.
- Affect the organization and plasticity of the genome:
  - Constitute an important part of it.
  - Create mutations (insertion inactivation, shunt mutations).
  - Increase the frequency of mutations (drosophila is estimated to induce up to half of spontaneous mutations).
  - Chromosome rearrangements (deletion, inversion, duplication).
  - Interaction of genomic components (chromosomes, plasmids...).
  - Indication of the place of integration of the transposon.

TE and chromosome rearrangements

- Transposable element (TRN) can act as a sight for recombination resulting in chromosome rearrangements.

- The recombination results in chromosome :
  - (A) Deletion.
  - (B) Inversion.
  - (C) Insertion.
  - (D) translocation.

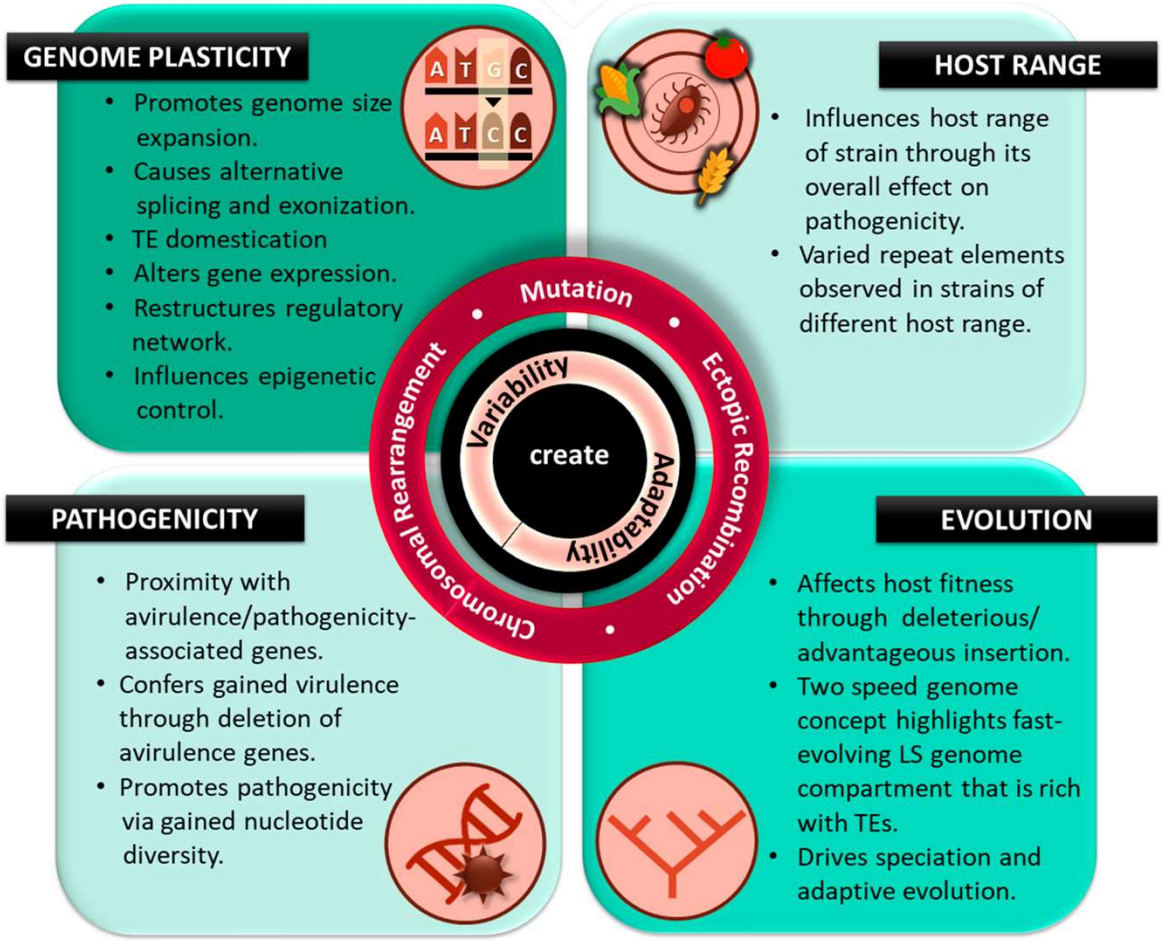


Ancestral DNA, Human Origins, and Migrations, 2018, Pages 61-103.

MUNI SCI

Impact of TE integration to the genome

- Phenotype changes.
- Insertion inactivation of the gene into which the transposon was incorporated (negative mutation).
- Acquisition of antibiotic resistance (positive mutation).
- Polar mutations affecting the expression of neighboring genes.



Int. J. Mol. Sci., 2019, 20(14), 3597.

MUNI SCI

THANK YOU FOR YOUR ATTENTION

- There is no such a thing in the nature that we can call junk. Just because of we don't get it...we don't have the right to call junk, thrash...etc.

https://slideplayer.com/slide/4784408/

http://biocomicals.blogspot.com/2011/08/junk-dna-version1.html

MUNI SCI