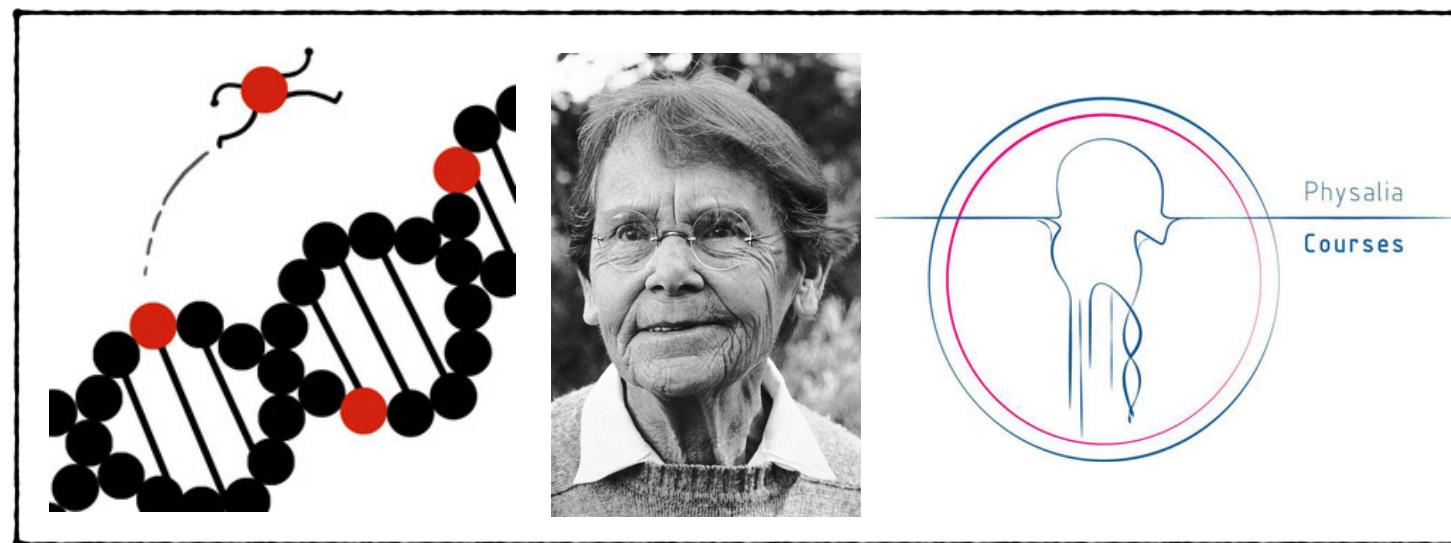


BIOINFORMATIC ANALYSIS OF TRANSPOSSABLE ELEMENTS

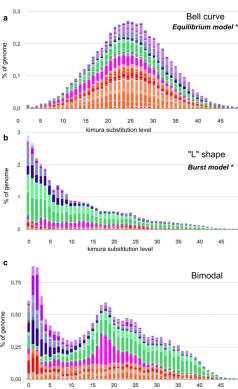
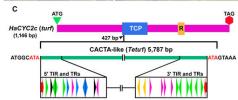
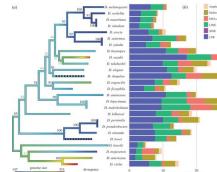
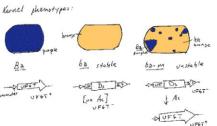
3rd-7th November 2025

Lecture 1

TE diversity and mechanisms

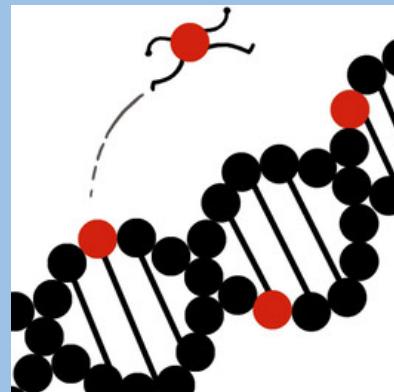


Valentina Peona

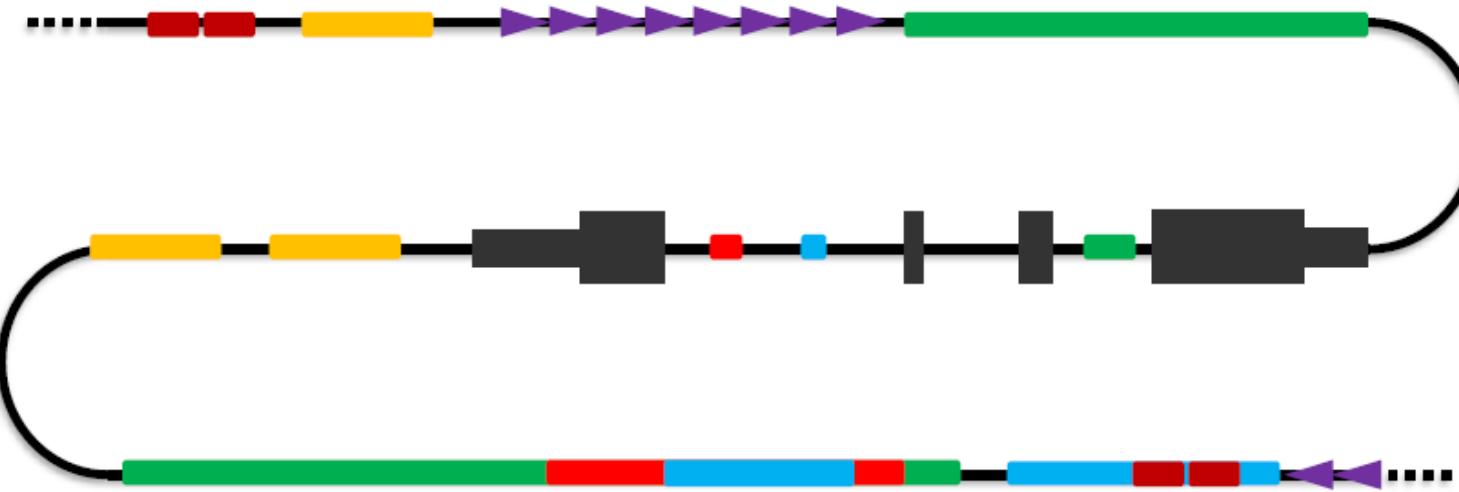


TE diversity

A FIELD GUIDE
TO THE TEs OF
THE WORLD



Genomes: DNA on repeats



Interspersed repeats

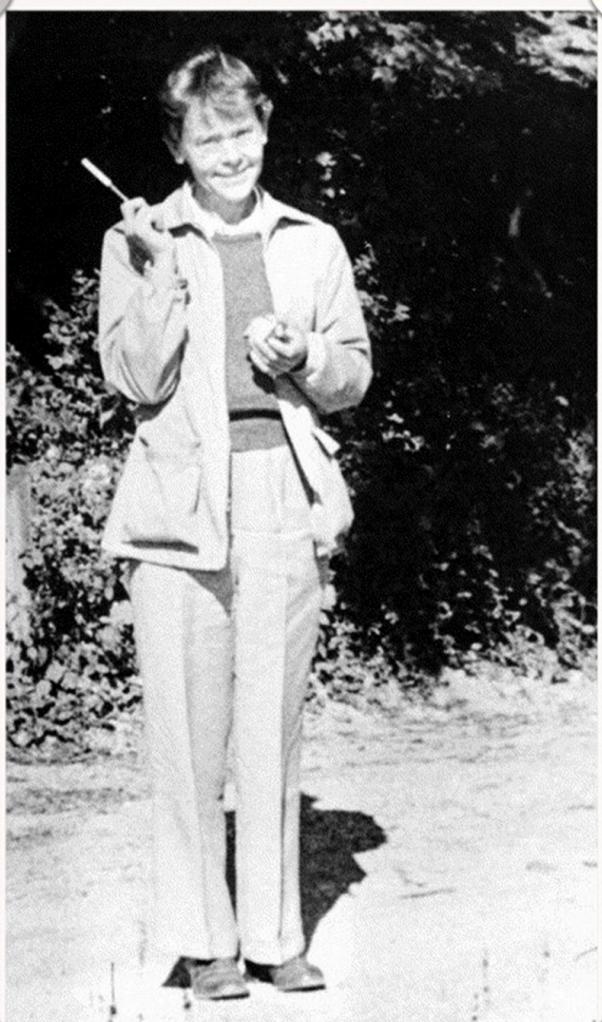
- Retrotransposons
- DNA transposons
- Endogenous viruses

Tandem repeats

- Satellites
- Minisatellites
- Microsatellites

Barbara McClintock

Controlling elements



FELDMAN, Jacqueline. "A Feeling for the Organism, The Life and Work of Barbara MacClintock." (1992): 430-433.

Feschotte, C. Transposable elements: McClintock's legacy revisited. *Nat Rev Genet* 24, 797–800 (2023). <https://doi.org/10.1038/s41576-023-00652-3>



Nobel Prize
1983

[Nobel Lecture](#)

**Barbara McClintock,
1951**

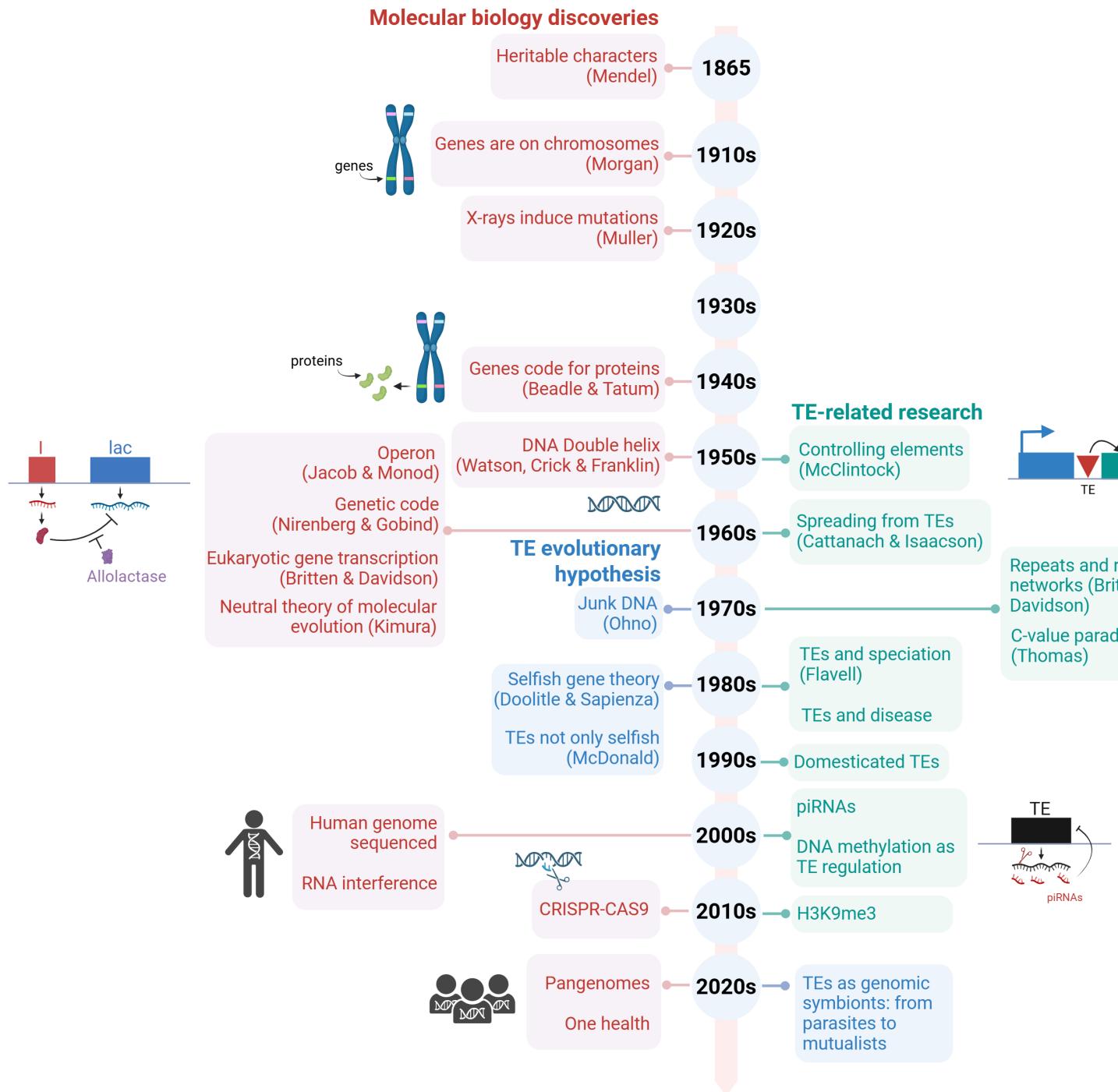


**Plant Breeding Department at Cornell
University in 1927**



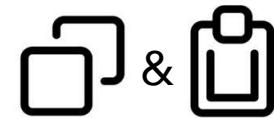
Standing from left to right are Charles R. Burnham, Marcus M. Rhoades, Rollins A. Emerson, and Barbara McClintock. Kneeling is George Beadle. Image courtesy of Cold Spring Harbor Laboratory Archives, NY. CC BY-NC-SA

TE History

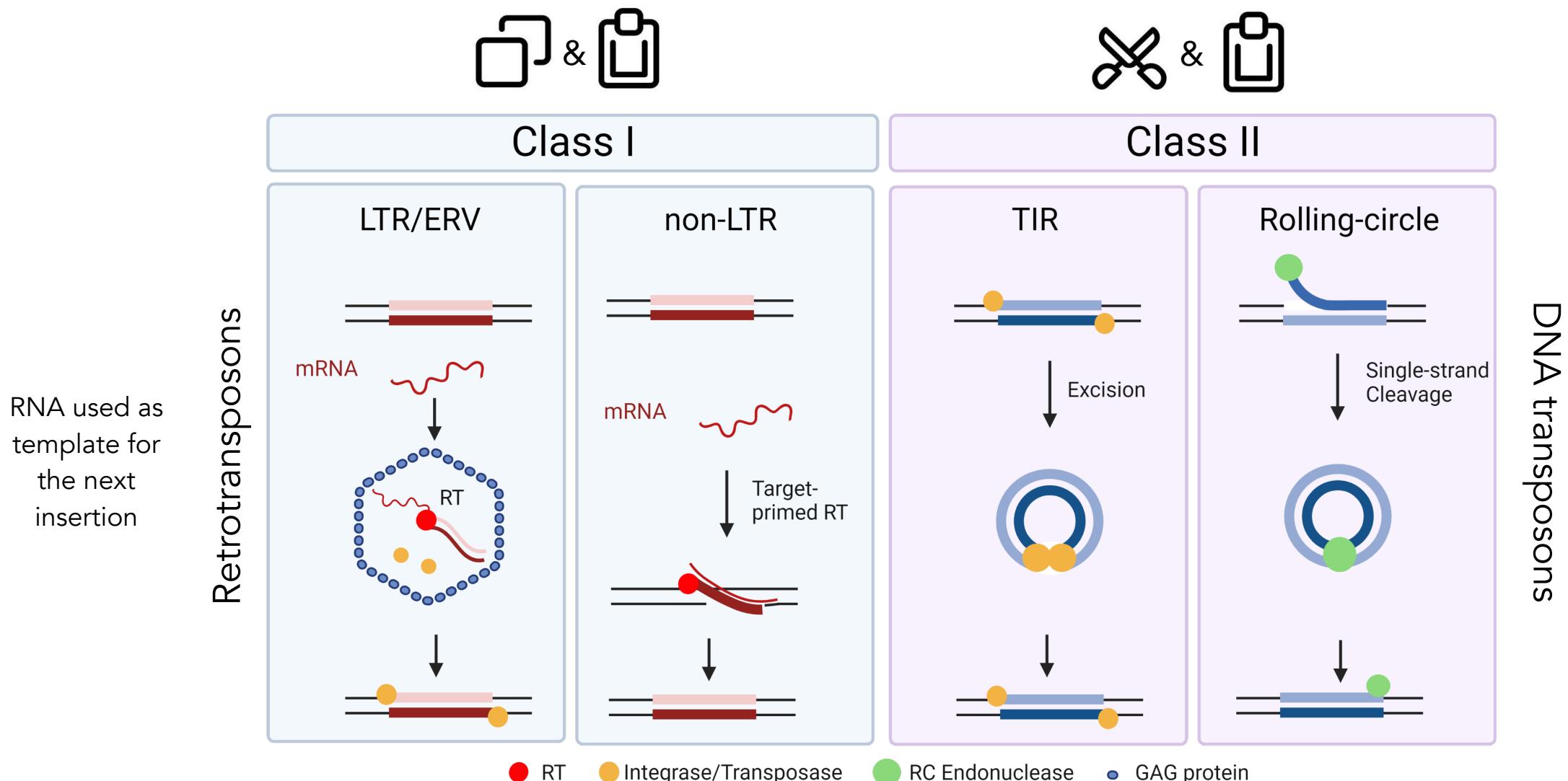


Eukaryotic transposable elements

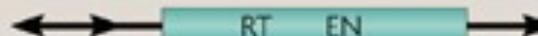
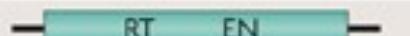
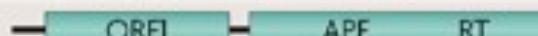
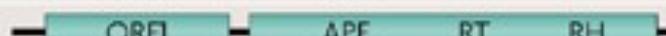
Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
Class I (retrotransposons)					
LTR	Copia	→ GAG AP INT RT RH →	4–6	RLC	P, M, F, O
	Gypsy	→ GAG AP RT RH INT →	4–6	RLG	P, M, F, O
	Bel-Pao	→ GAG AP RT RH INT →	4–6	RLB	M
	Retrovirus	→ GAG AP RT RH INT ENV →	4–6	RLR	M
	ERV	→ GAG AP RT 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	RITE	M
	Jockey	ORFI APE RT	Variable	RIJ	M
	L1	ORFI APE RT	Variable	RIL	P, M, F, O
	I	ORFI 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
Class II (DNA transposons) - Subclass 1					
TIR	Tc1-Mariner	→ Tase* →	TA	DTT	P, M, F, O
	hAT	→ Tase* →	8	DTA	P, M, F, O
	Mutator	→ Tase* →	9–11	DTM	P, M, F, O
	Merlin	→ Tase* →	8–9	DTE	M, O
	Transib	→ Tase* →	5	DTR	M, F
	P	→ Tase →	8	DTP	P, M
	PiggyBac	→ Tase →	TTAA	DTB	M, O
	PIF-Harbinger	→ Tase* ORF2 →	3	DTH	P, M, F, O
	CACTA	→ → Tase → ORF2 → →	2–3	DTC	P, M, F
	Crypton	YR	0	DYC	F
Class II (DNA transposons) - Subclass 2					
Helitron	Helitron	— RPA — Y2 HEL —	0	DHH	P, M, F
Maverick	Maverick	— C-INT ATP — CYP POL B —	6	DMM	M, F, O
Structural features 					
Protein coding domains 					
AP, Aspartic protease APE, Apurinic endonuclease ATP, Packaging ATPase C-INT, C-integrase ENV, Envelope protein GAG, Capsid protein HEL, Helicase INT, Integrase PCL B, DNA polymerase B RH, RNase H RPA, Replication protein A (found only in plants) Tase, Transposase (* with DDE motif) YR, Tyrosine recombinase CYP, Cysteine protease EN, Endonuclease ORF, Open reading frame of unknown function RT, Reverse transcriptase Y2, YR with YY motif					
Species groups P, Plants M, Metazoans F, Fungi O, Others					



Eukaryotic transposable elements



Class I: LINE retrotransposons

Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
Class I (retrotransposons)					
PLE	Penelope		Variable	RPP	P, M, F, O
LINE	R2		Variable	RIR	M
	RTE		Variable	RIT	M
	Jockey		Variable	RIJ	M
	L1		Variable	RIL	P, M, F, O
	I		Variable	RII	P, M, F

RT - retrotranscriptase

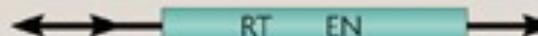
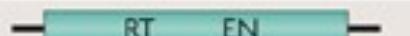
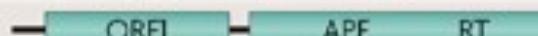
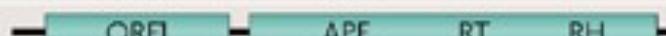
EN - endonuclease

RH – RNase H

APE - DNA (apurinic/apyrimidinic site) endonuclease



Class I: LINE retrotransposons

Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
Class I (retrotransposons)					
PLE	Penelope		Variable	RPP	P, M, F, O
LINE	R2		Variable	RIR	M
	RTE		Variable	RIT	M
	Jockey		Variable	RIJ	M
	L1		Variable	RIL	P, M, F, O
	I		Variable	RII	P, M, F

RT - retrotranscriptase

EN - endonuclease

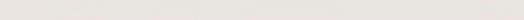
RH – RNase H

APE - DNA (apurinic/apyrimidinic site) endonuclease



Promoter for
RNA polymerase

Class I: LINE retrotransposons

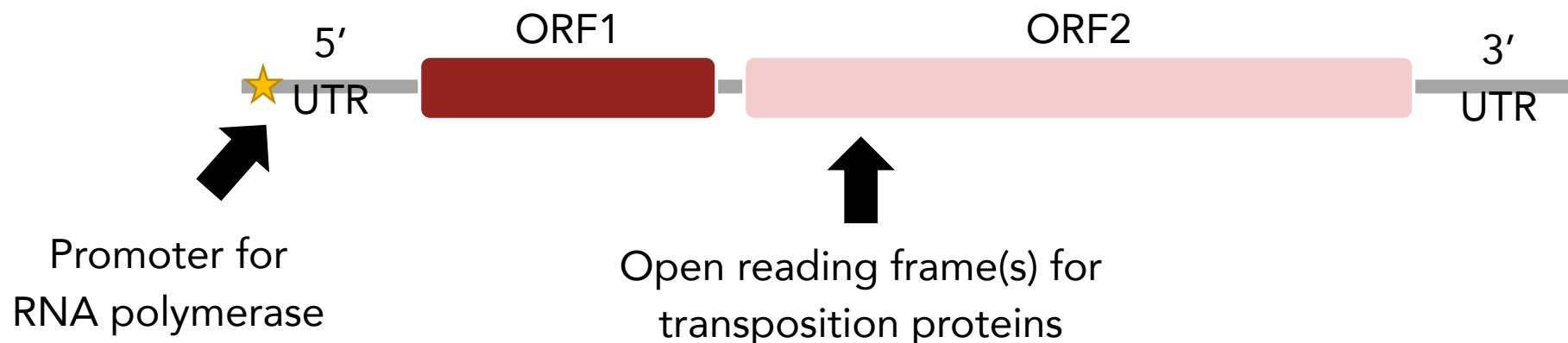
Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
Class I (retrotransposons)					
PLE	Penelope		Variable	RPP	P,M,F,O
LINE	R2		Variable	RIR	M
	RTE		Variable	RIT	M
	Jockey		Variable	RIJ	M
	L1		Variable	RIL	P,M,F,O
	I		Variable	RII	P,M,F

RT - retrotranscriptase

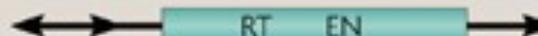
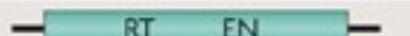
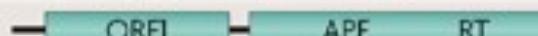
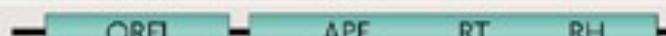
EN - endonuclease

RH – RNase H

APE - DNA (apurinic/apyrimidinic site) endonuclease



Class I: LINE retrotransposons

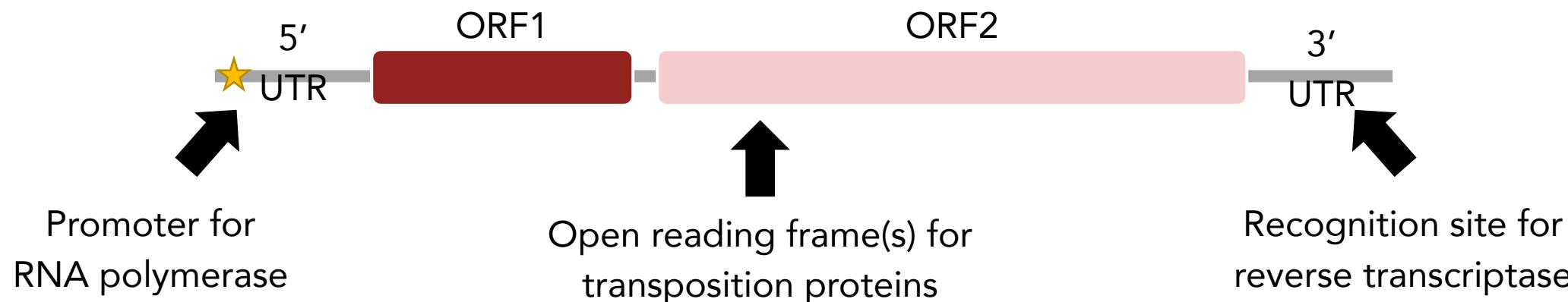
Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
Class I (retrotransposons)					
PLE	Penelope		Variable	RPP	P, M, F, O
LINE	R2		Variable	RIR	M
	RTE		Variable	RIT	M
	Jockey		Variable	RIJ	M
	L1		Variable	RIL	P, M, F, O
	I		Variable	RII	P, M, F

RT - retrotranscriptase

EN - endonuclease

RH – RNase H

APE - DNA (apurinic/apyrimidinic site) endonuclease



Class I: LINE retrotransposons

Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
Class I (retrotransposons)					
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

RT - retrotranscriptase

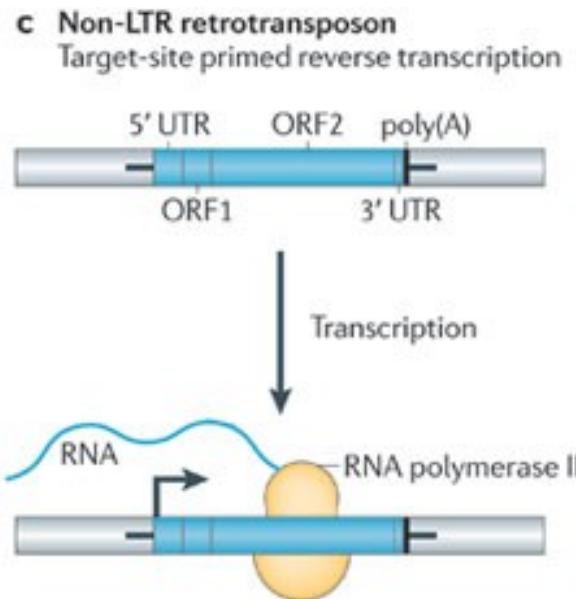
EN - endonuclease

RH – RNase H

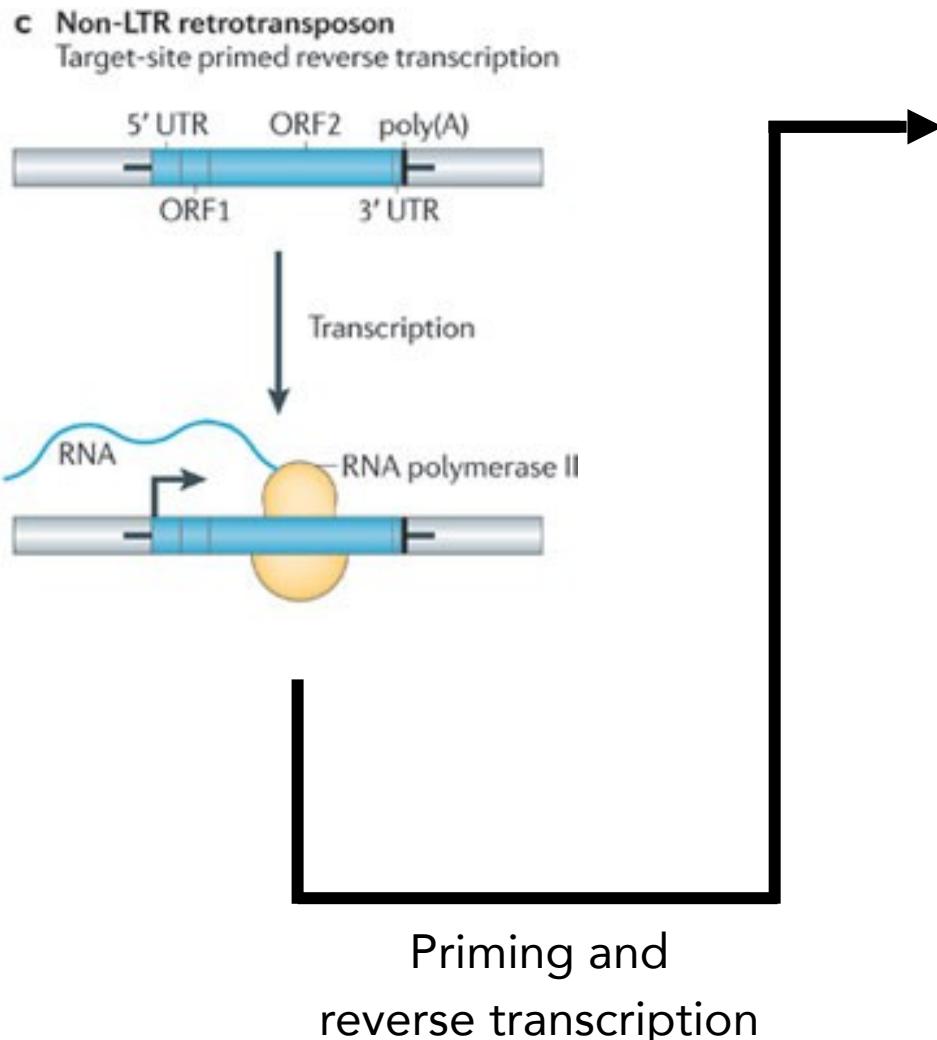
APE - DNA (apurinic/apyrimidinic site) endonuclease



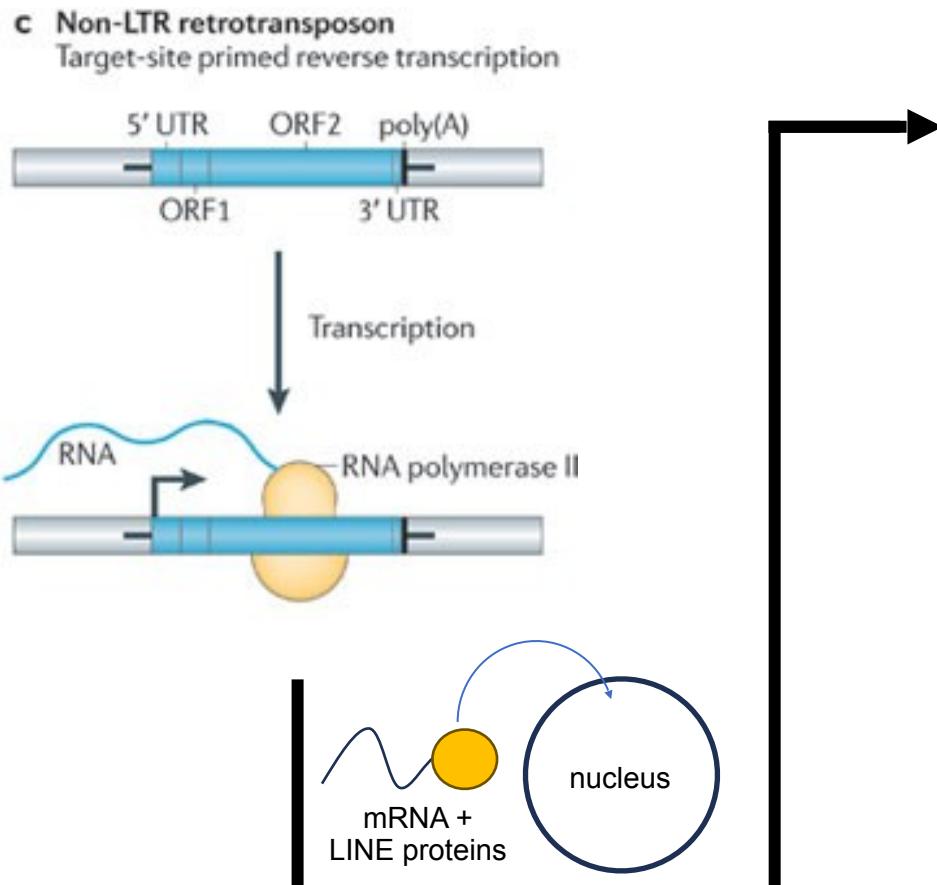
Target-primed reverse transcription (TPRT)



Target-primed reverse transcription (TPRT)



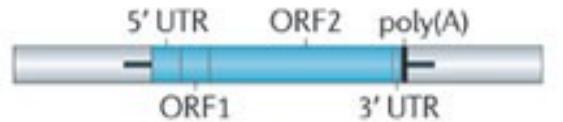
Target-primed reverse transcription (TPRT)



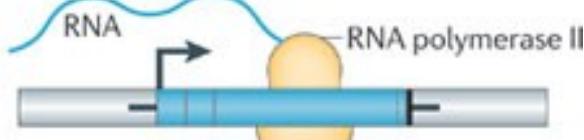
Priming and
reverse transcription

Target-primed reverse transcription (TPRT)

c Non-LTR retrotransposon
Target-site primed reverse transcription

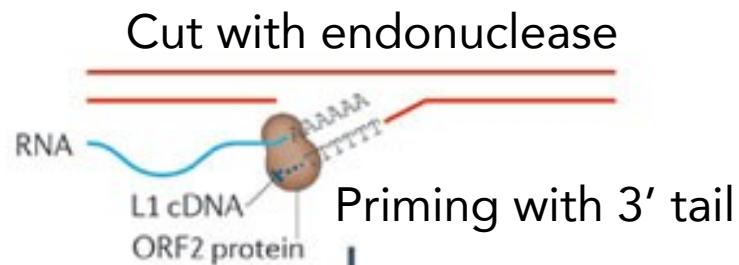


Transcription

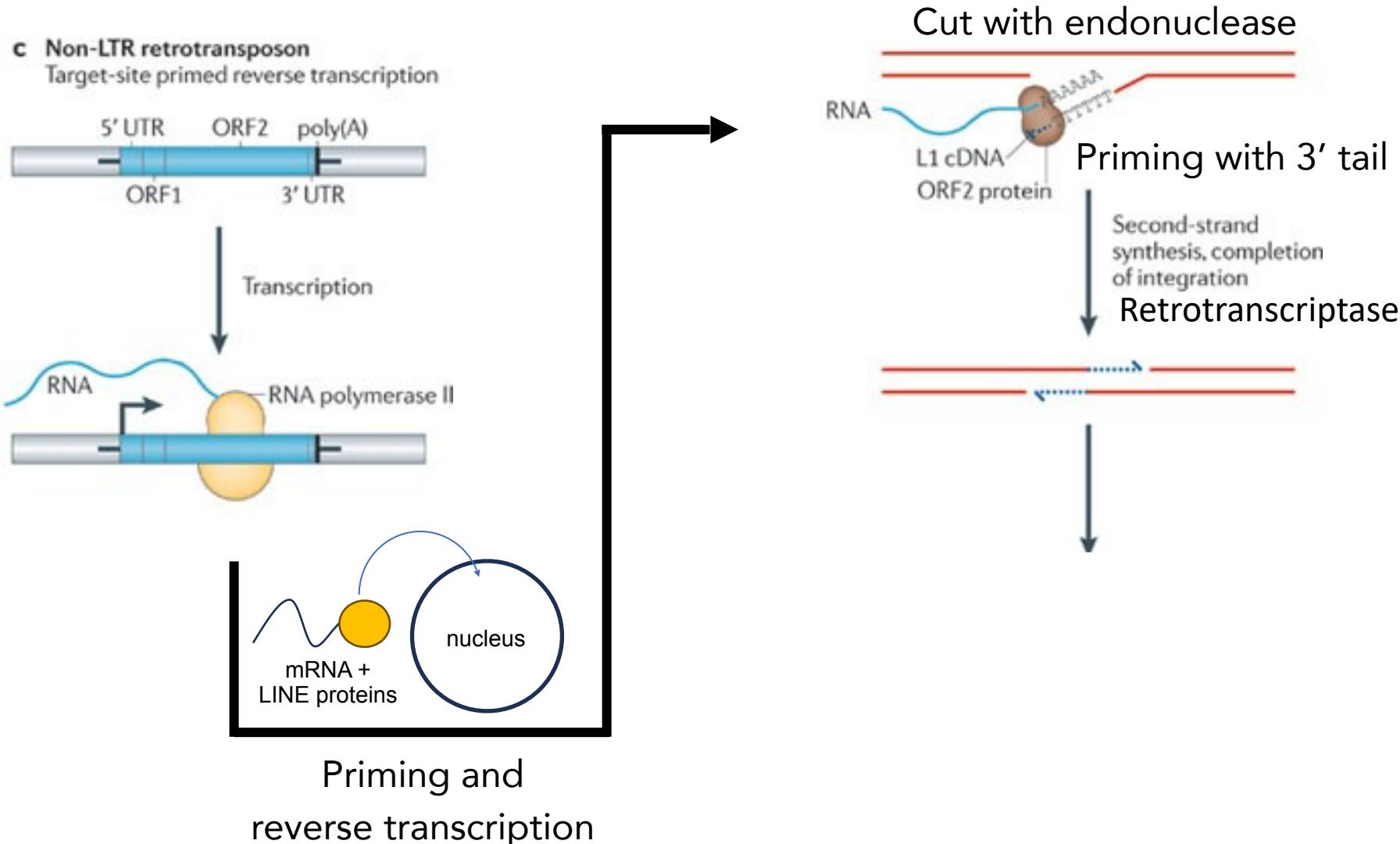


mRNA +
LINE proteins

Priming and
reverse transcription



Target-primed reverse transcription (TPRT)

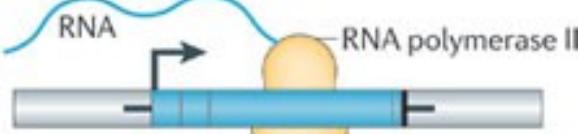


Target-primed reverse transcription (TPRT)

c Non-LTR retrotransposon
Target-site primed reverse transcription

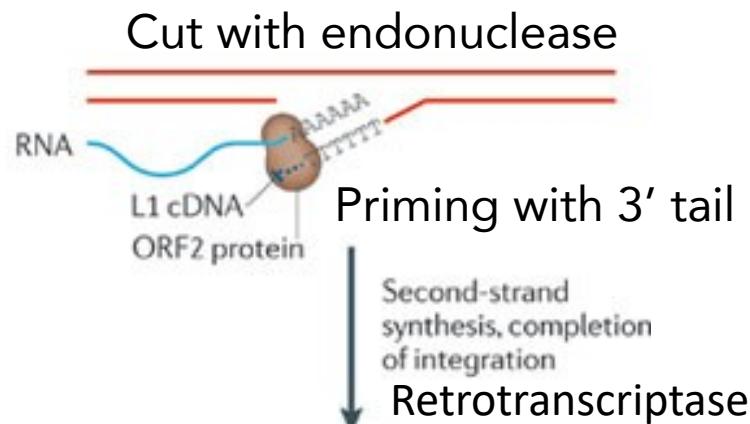


Transcription



mRNA +
LINE proteins

Priming and
reverse transcription



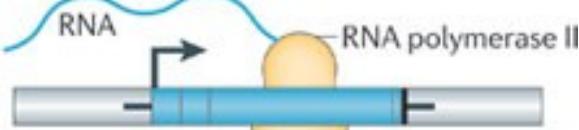
Nature Reviews | Genetics

Target-primed reverse transcription (TPRT)

c Non-LTR retrotransposon
Target-site primed reverse transcription



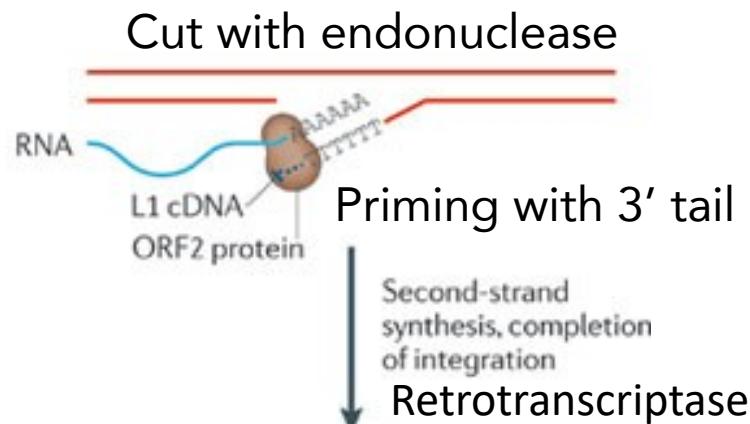
Transcription



mRNA +
LINE proteins

nucleus

Priming and
reverse transcription



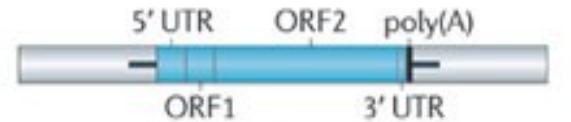
Nature Reviews | Genetics

TPRT often undergoes premature 5' truncation and loss of promoters and/or protein

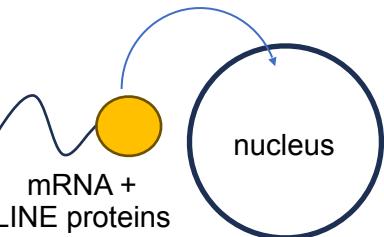
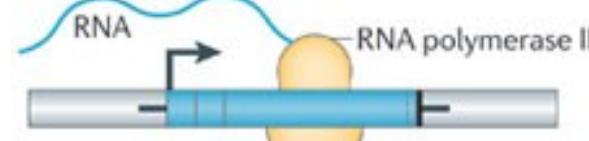
Levin and Moran 2011, Nat Rev Gen

Target-primed reverse transcription (TPRT)

c Non-LTR retrotransposon
Target-site primed reverse transcription



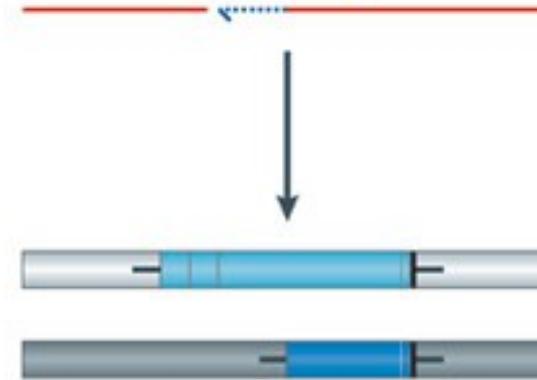
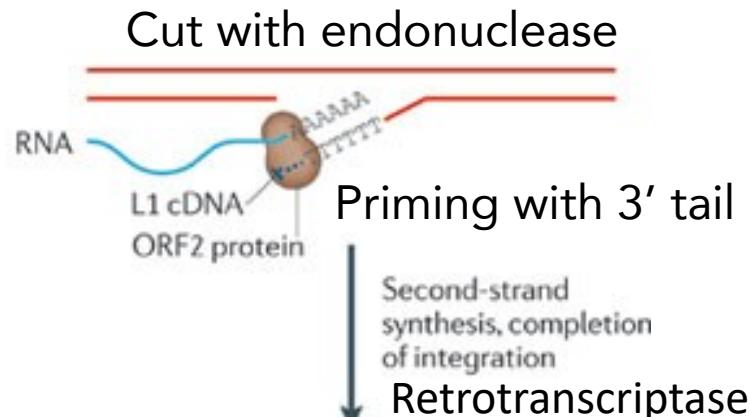
Transcription



UPDATE

George E. Ghanim et al., Structural mechanism of LINE-1 target-primed reverse transcription. *Science* 388, eads8412 (2025).

Priming and
reverse transcription

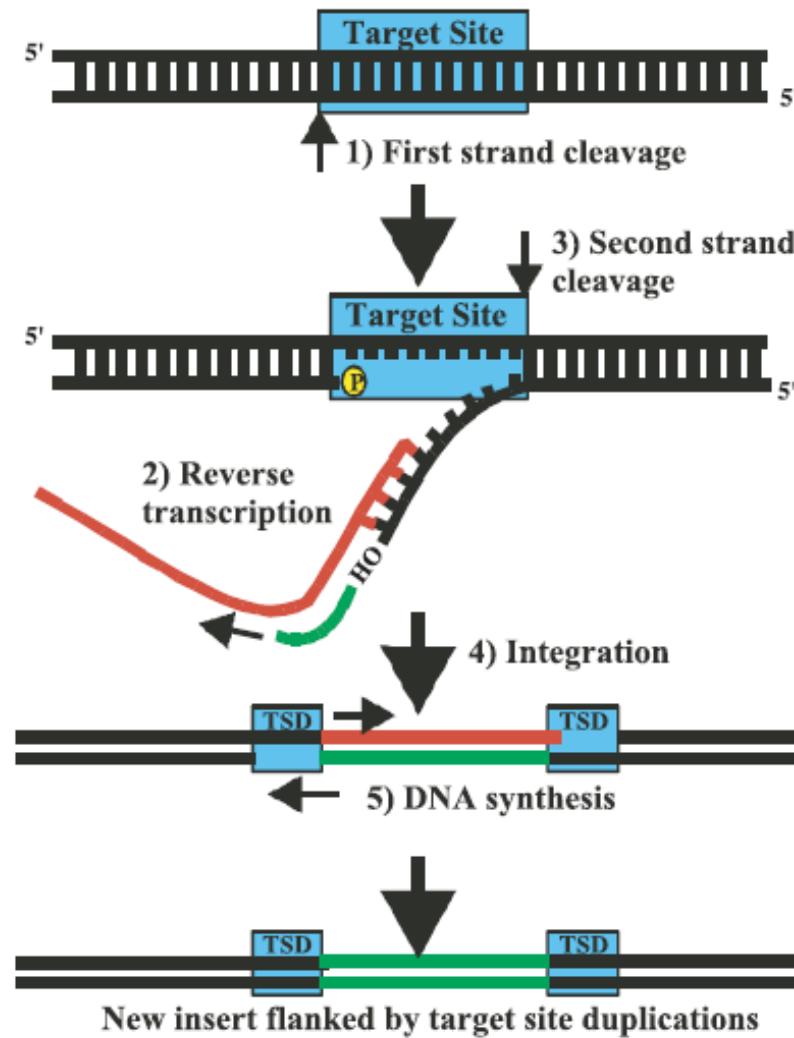


Nature Reviews | Genetics

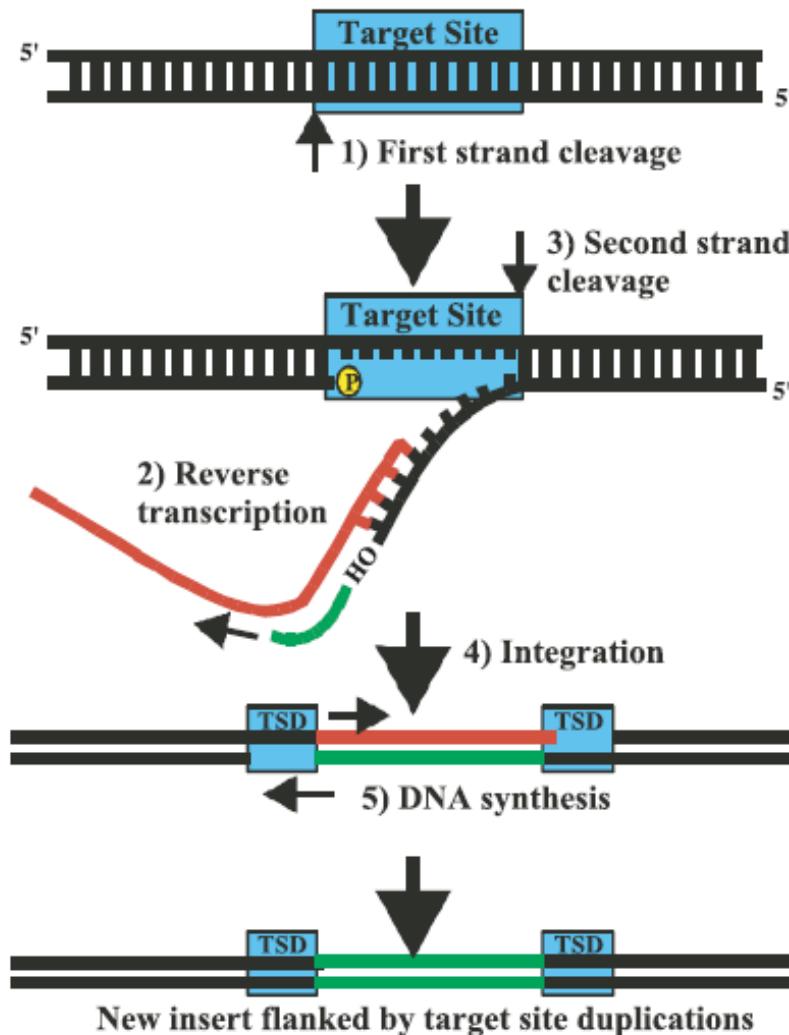
TPRT often undergoes premature 5' truncation and loss of promoters and/or protein

Levin and Moran 2011, Nat Rev Gen

Target Site Duplication (TSD)

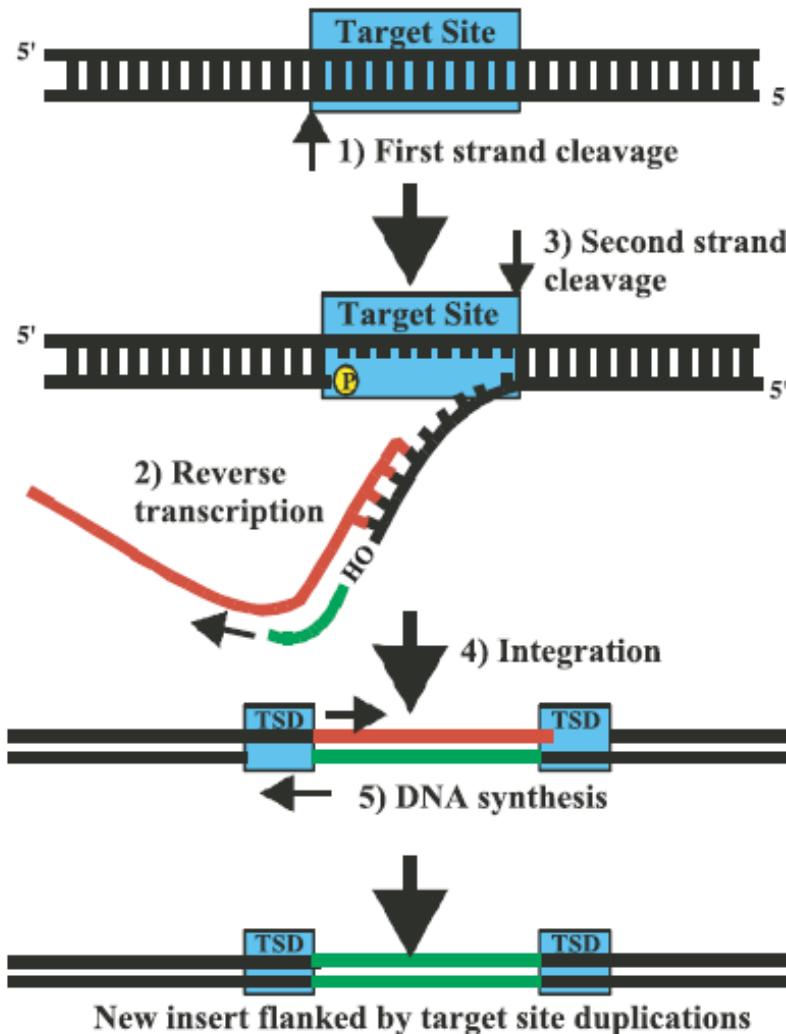


Target Site Duplication (TSD)



TSDs are the hallmark of most (retro)transposons

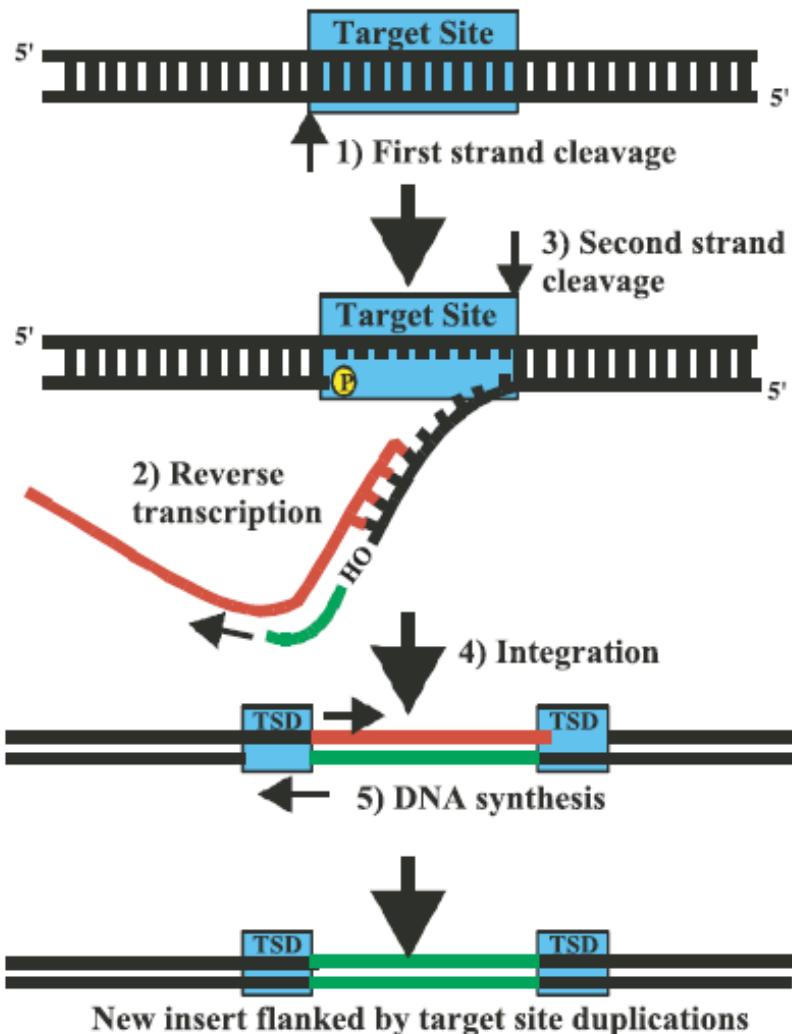
Target Site Duplication (TSD)



The length is variable for LINEs
but can be of specific lengths
for other types of elements

TSDs are the hallmark of most (retro)transposons

Target Site Duplication (TSD)

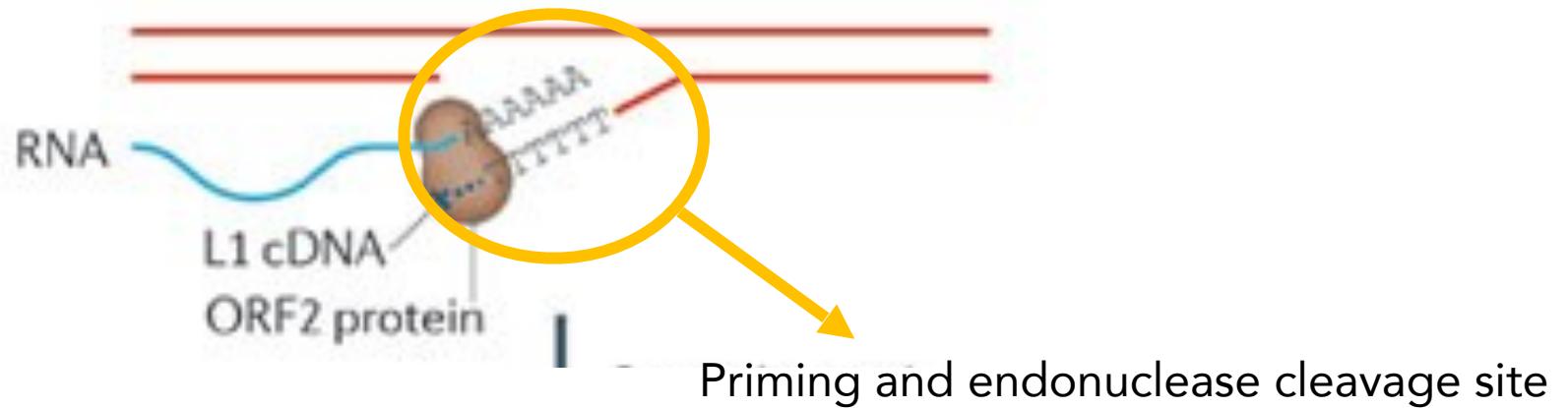


The length of TSDs is important for classification

The length is variable for LINEs but can be of specific lengths for other types of elements

TSDs are the hallmark of most (retro)transposons

Target site preferentiality



Preferential target for L1 in
humans:
5'-TTTT/AA-3'

Target site preferentiality

The case of R2

R2 target site

Element name	3'-terminal sequences of R2 elements	28S rDNA
R2Eb (hagfish)	AAAGTAGTATGTGCTCATGACCAAGACCAGGTCTGCAGGGACAACATATCA	TAGCCAAATGC
R2Ta (minnow)	ACCGTTCCAGGAACGGTATCAGCTCAGACTCGGTGGTATGGCCTCACAG	TAGCCAAATGC
R201-A (medaka fish)	CCCCGTGACGGACACATTAACTCCGGAAAGCGAGTGGTGACTCGCCTCAAG	TAGCCAAATGC
R2Cr-A (turtle)	TTGGTACACCCCAAATAACTCGAACAGAAAACCGCGTGGGTTGTTAAA	TAGCCAAATGC
R2Cr-B1 (turtle)	TTTATTCACTCCGAATAACTCGAAAAAGAAAACCCATGTGAGTTTTAAAAA	TAGCCAAATGC
R2Cr-B2 (turtle)	GGTGTACACTCCGAATAACCCGAAAAAGAAAACCCGCGAGGGTTTTAAAAA	TAGCCAAATGC
R2Mr (sea lily)	CGAGCGTGCCACCTTTGCACCGTTCCGTGCCGATGGGATTAGGCGAAA	TAGCCAAATGC
R2Ha (jumper spider)	CACTACGGTACCCGCCGTATTGGGGAAATGCGCTTCGGCGTAGAGTT	TAGCCAAATGC
R2Pc (crayfish)	TAAAAAACCTATACCCCCCCAGAAGTGGCAAGGGGAGTTAGGGACACA ₁₄	TAGCCAAATGC
R2T1 (tadpole shrimp)	GGTCTTACTTGCAGGCCACCCAAAAATGCAAGAAAGTTTTATTAATAAAA	TAGCCAAATGC
R2Amel (honeybee)	AGGTATCCTCGAAATTCCGCCACGATCCTCTGATCGTAGGCGAAAACA	TAGCCAAATGC
R2Sm-A (bloodfluke)	GCCAAGCCCATTGGGTCTGTACAATTGATACTAAAAATAATGTTAT	TAGCCAAATGC

Target site preferentiality

The case of R2

R2 target site

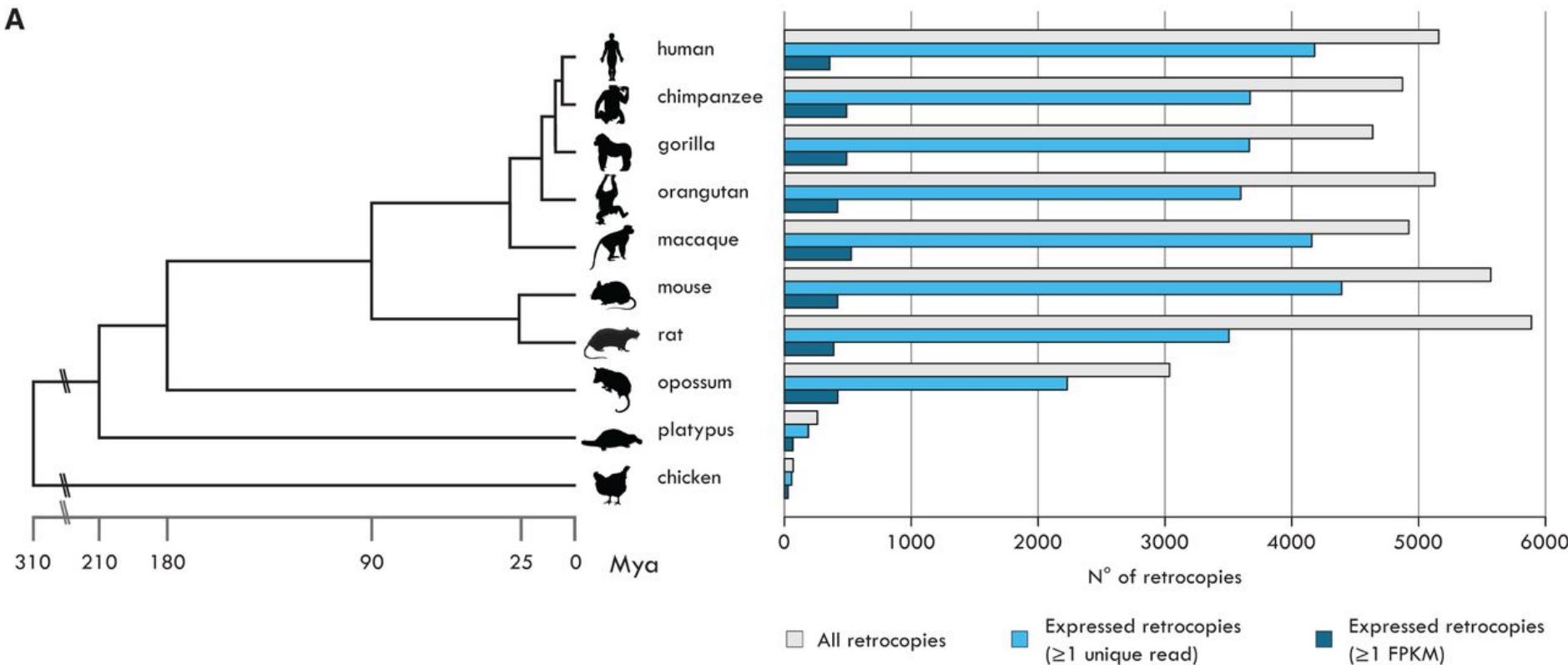
Element name	3'-terminal sequences of R2 elements	28S rDNA
R2Eb (hagfish)	AAAGTAGTATGTGCTCATGACCAAGACCAGGTCTGCAGGGACAACATATCA	TAGCCAAATGC
R2Ta (minnow)	ACCGTTCCAGGAACGGTATCAGCTCAGACTCGGTGGTATGCCCTCACAG	TAGCCAAATGC
R201-A (medaka fish)	CCCCGTGACGGACACATTAACTCCGGAAAGCGAGTGGTGACTCGCCTCAAG	TAGCCAAATGC
R2Cr-A (turtle)	TTGGTACACCCCAAATAACTCGAACAGAAAACCGCGTGGGTTGTTAAA	TAGCCAAATGC
R2Cr-B1 (turtle)	TTTATTCACTCCGAATAACTCGAAAAAGAAAACCCATGTGAGTTTTAAAAA	TAGCCAAATGC
R2Cr-B2 (turtle)	GGTGTACACTCCGAATAACCCGAAAAAGAAAACCGCGAGGGTTTTAAAAA	TAGCCAAATGC
R2Mr (sea lily)	CGAGCGTGCCACCTTTGCACCGTTCCGTGCCGATGGGATTAGGCGAAA	TAGCCAAATGC
R2Ha (jumper spider)	CACTACGGTACCCGCCGTATTGGGGAAATGCGCTTCGGCGTAGAGTT	TAGCCAAATGC
R2Pc (crayfish)	TAAAAAACCTATACCCCCCGAAGTGGCAAGGGGAGTTAGGGACACA ₁₄	TAGCCAAATGC
R2T1 (tadpole shrimp)	GGTCTTACTTGCAGGCCACCCAAAAATGCAAGAAAGTTTTATTAATAAAA	TAGCCAAATGC
R2Amel (honeybee)	AGGTATCCTCGAAATTCCGCCACGATCCTCTGATCGTAGGCGAAAACA	TAGCCAAATGC
R2Sm-A (bloodfluke)	GCCAAGCCCATTGGGTCTGTACAATTGATACTAAAAATAATGTTAT	TAGCCAAATGC

TPRT can lead to insertion site preference or even specificity

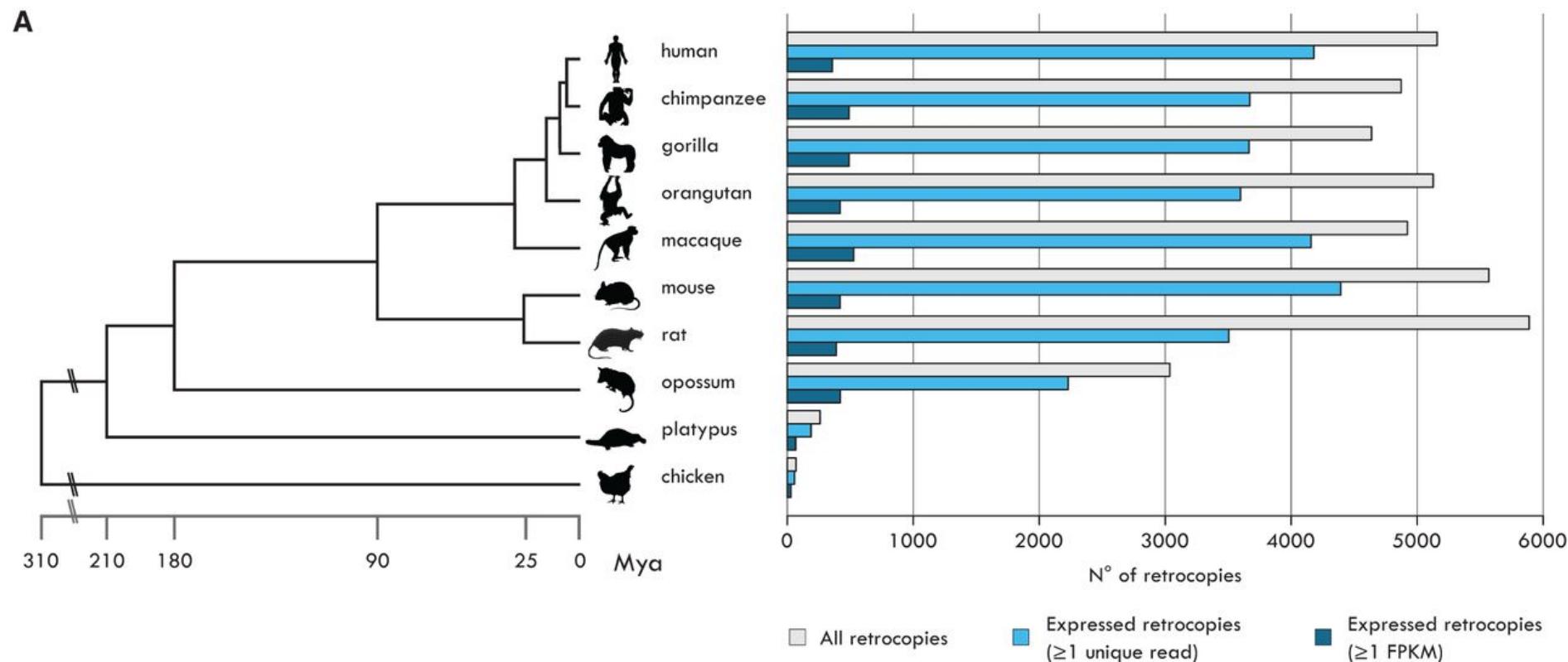
L1 and retro(pseudo)genes



A



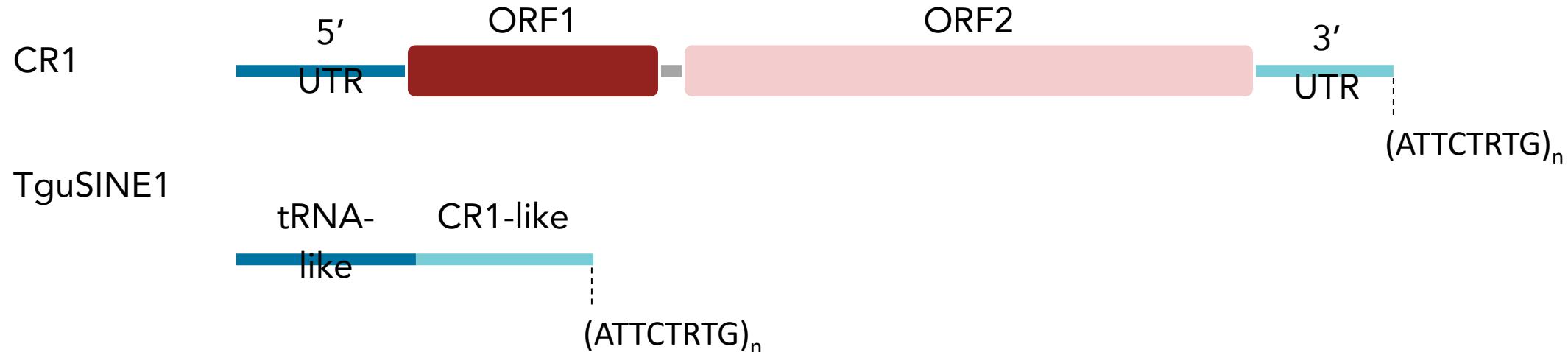
L1 and retro(pseudo)genes



Retrogenes occur when LINE RT recognizes the poly-A tails (L1)

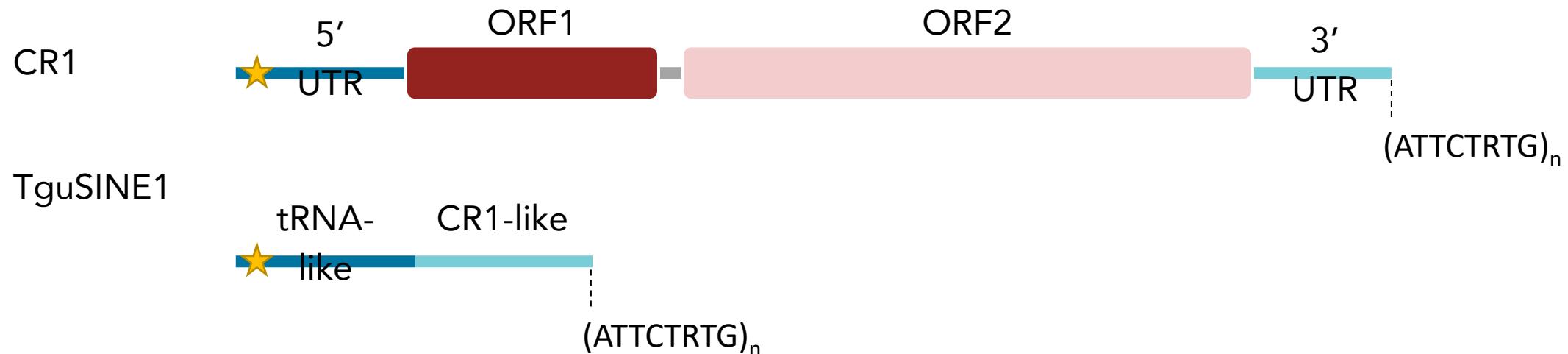
Class I: SINE retrotransposons

Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
<i>Class I (retrotransposons)</i>					
SINE	tRNA		Variable	RST	P, M, F
	7SL		Variable	RSL	P, M, F
	5S		Variable	RSS	M, O



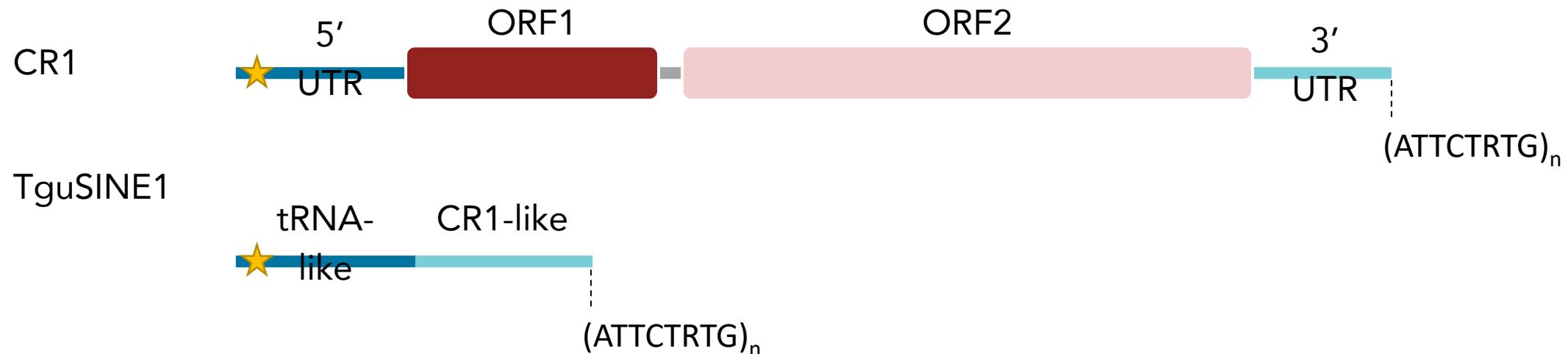
Class I: SINE retrotransposons

Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
<i>Class I (retrotransposons)</i>					
SINE	tRNA		Variable	RST	P, M, F
	7SL		Variable	RSL	P, M, F
	5S		Variable	RSS	M, O



Class I: SINE retrotransposons

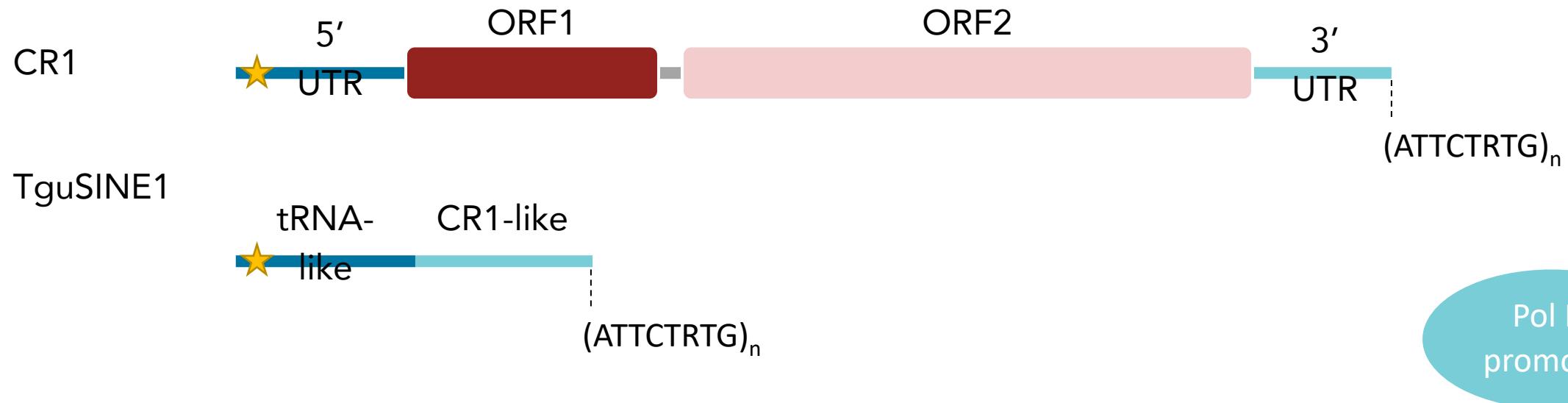
Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
<i>Class I (retrotransposons)</i>					
SINE	tRNA		Variable	RST	P, M, F
	7SL		Variable	RSL	P, M, F
	5S		Variable	RSS	M, O



SINEs use the LINE protein machinery to move and replicate – trans-mobilization!
Non-autonomous elements

Class I: SINE retrotransposons

Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
Class I (retrotransposons)					
SINE	tRNA		Variable	RST	P, M, F
	7SL		Variable	RSL	P, M, F
	5S		Variable	RSS	M, O



SINEs use the LINE protein machinery to move and replicate – trans-mobilization!
Non-autonomous elements

TE characteristics so far...

- Promoter is necessary
- Encode proteins and recognition motifs for their own mobilization
- Transpositions leave signatures in the genome: target site duplications
- Their particular mobilization can have different effects on genome evolution:
 - clustering (target site preferentiality)
 - trans-mobilization of other mRNAs (e.g., non-autonomous TEs and retrogenes)

Class I: LTR 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

GAG – capsid protein

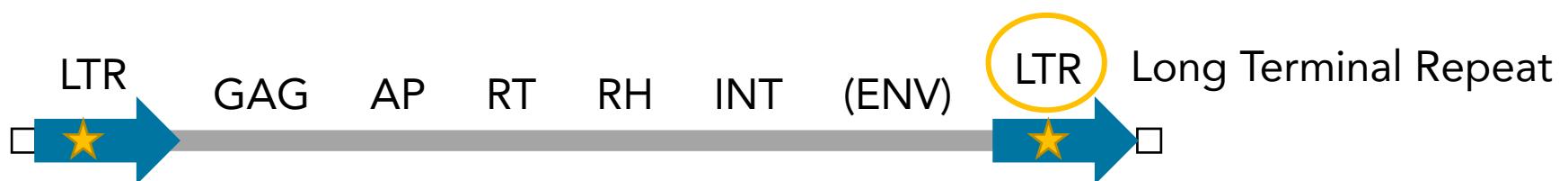
AP – aspartic proteinase

RT - retrotranscriptase

RH – RNase H

INT - integrase

ENV – envelope protein



Class I: LTR 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

GAG – capsid protein

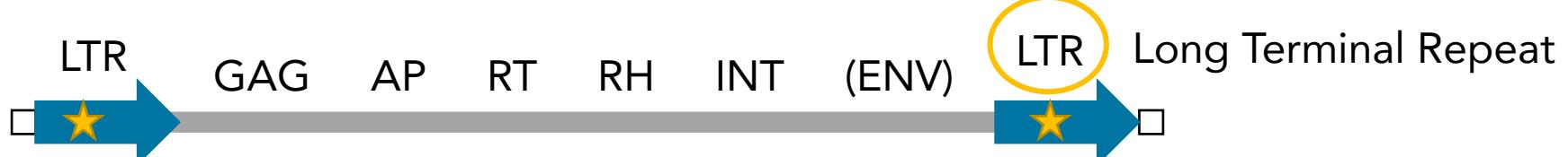
AP – aspartic proteinase

RT - retrotranscriptase

RH – RNase H

INT - integrase

ENV – envelope protein



Non-allelic homologous recombination (NAHR)

Full-length



Class I: LTR 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

GAG – capsid protein

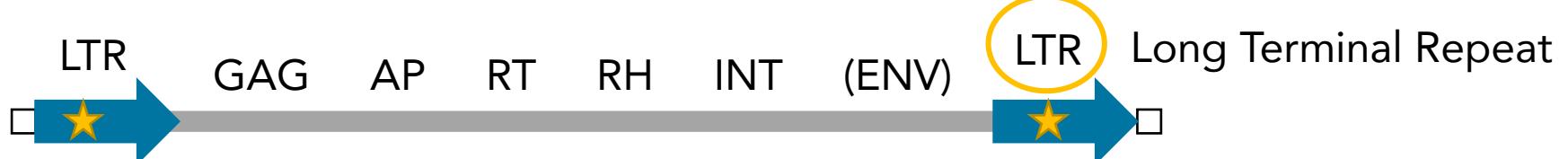
AP – aspartic proteinase

RT - retrotranscriptase

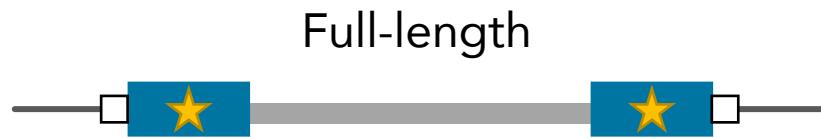
RH – RNase H

INT - integrase

ENV – envelope protein



Non-allelic homologous recombination (NAHR)



Class I: LTR 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

GAG – capsid protein

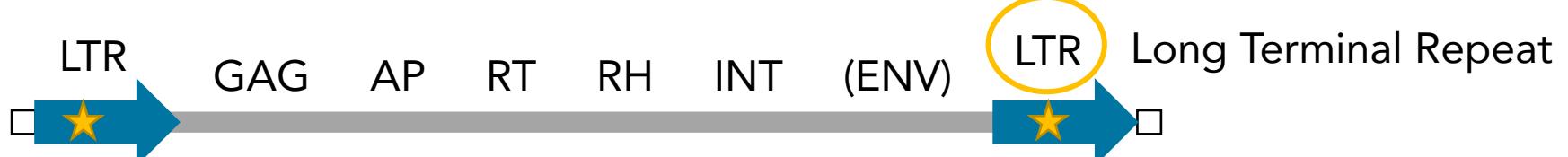
AP – aspartic proteinase

RT - retrotranscriptase

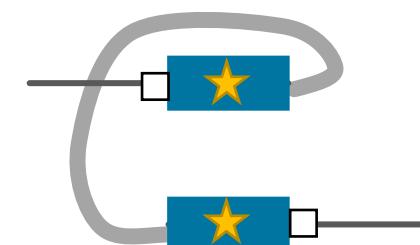
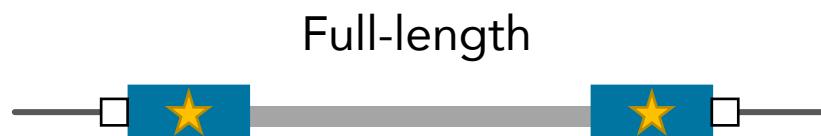
RH – RNase H

INT - integrase

ENV – envelope protein

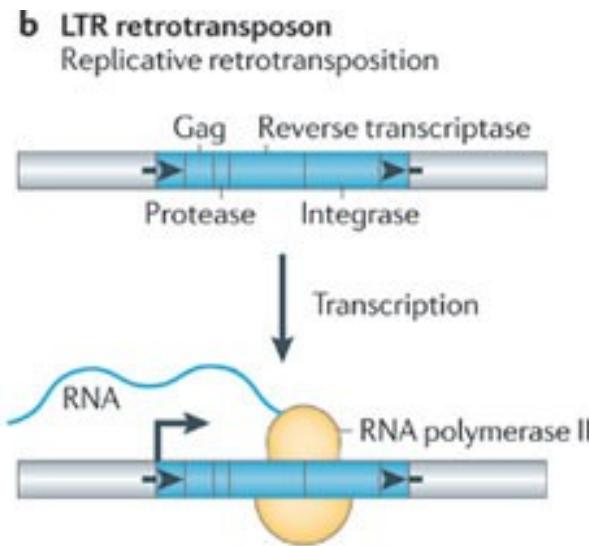


Non-allelic homologous recombination (NAHR)

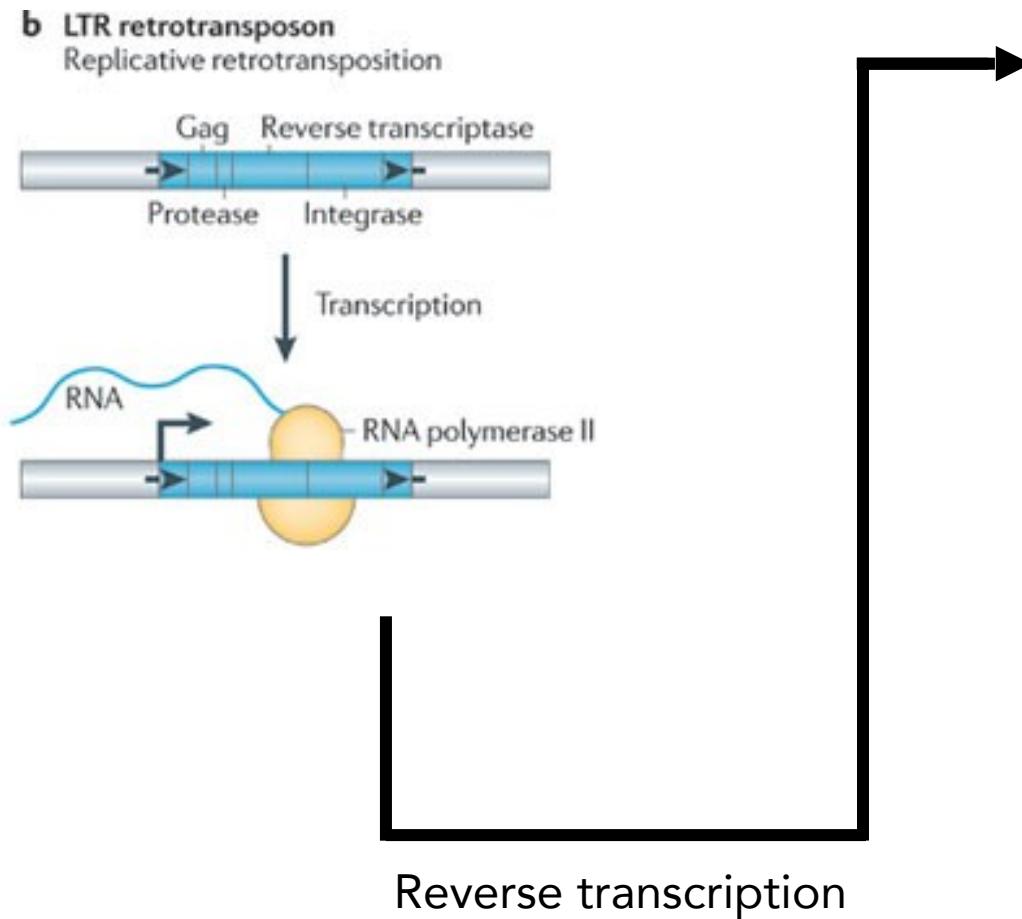


Wicker et al 2007, Nat Rev Gen

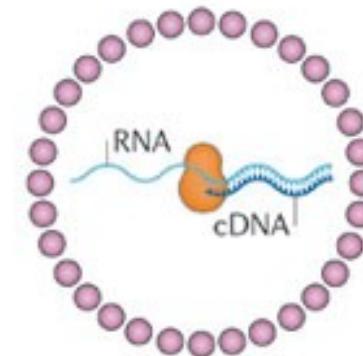
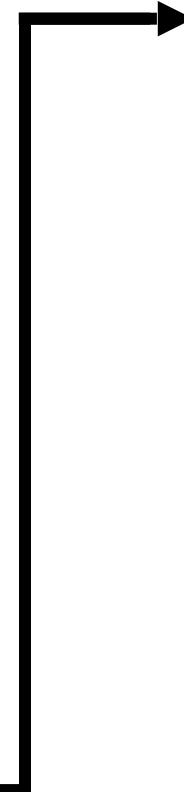
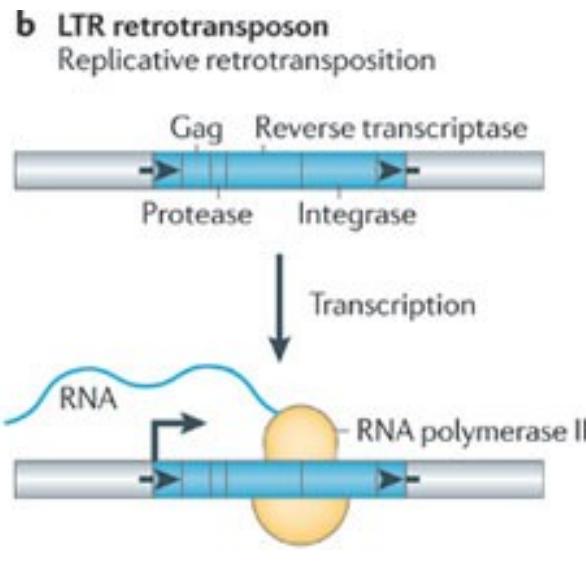
Replicative retrotransposition



Replicative retrotransposition

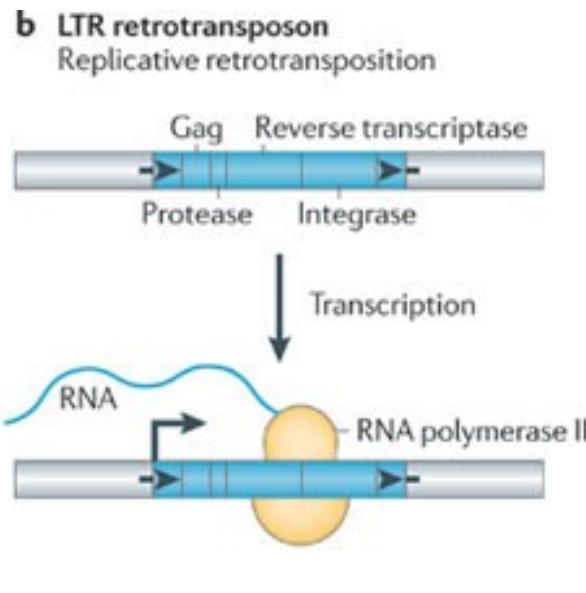


Replicative retrotransposition

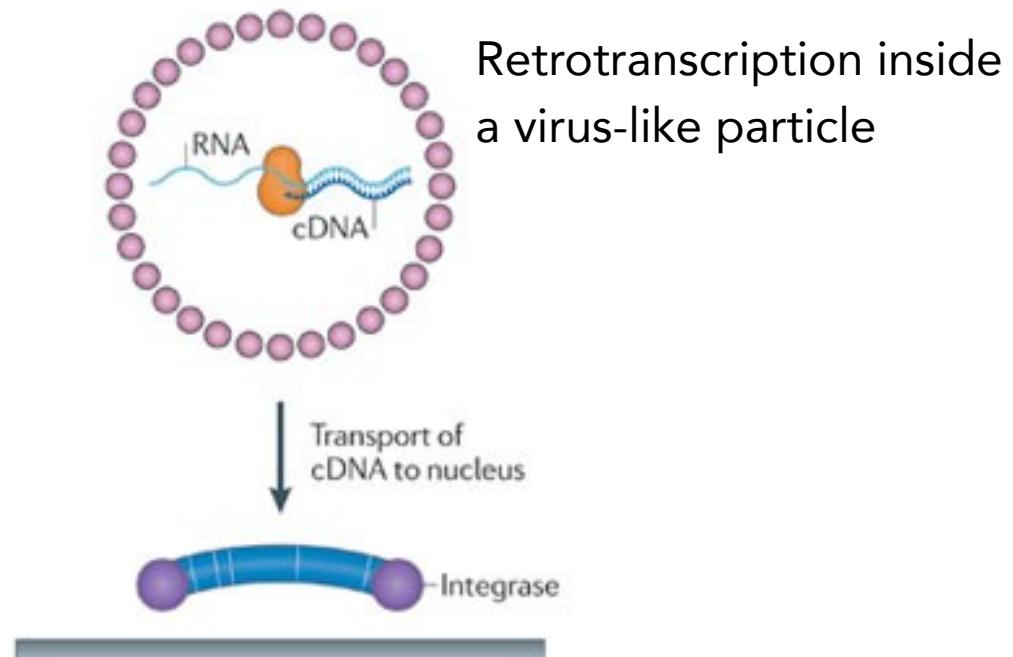


Retrotranscription inside
a virus-like particle

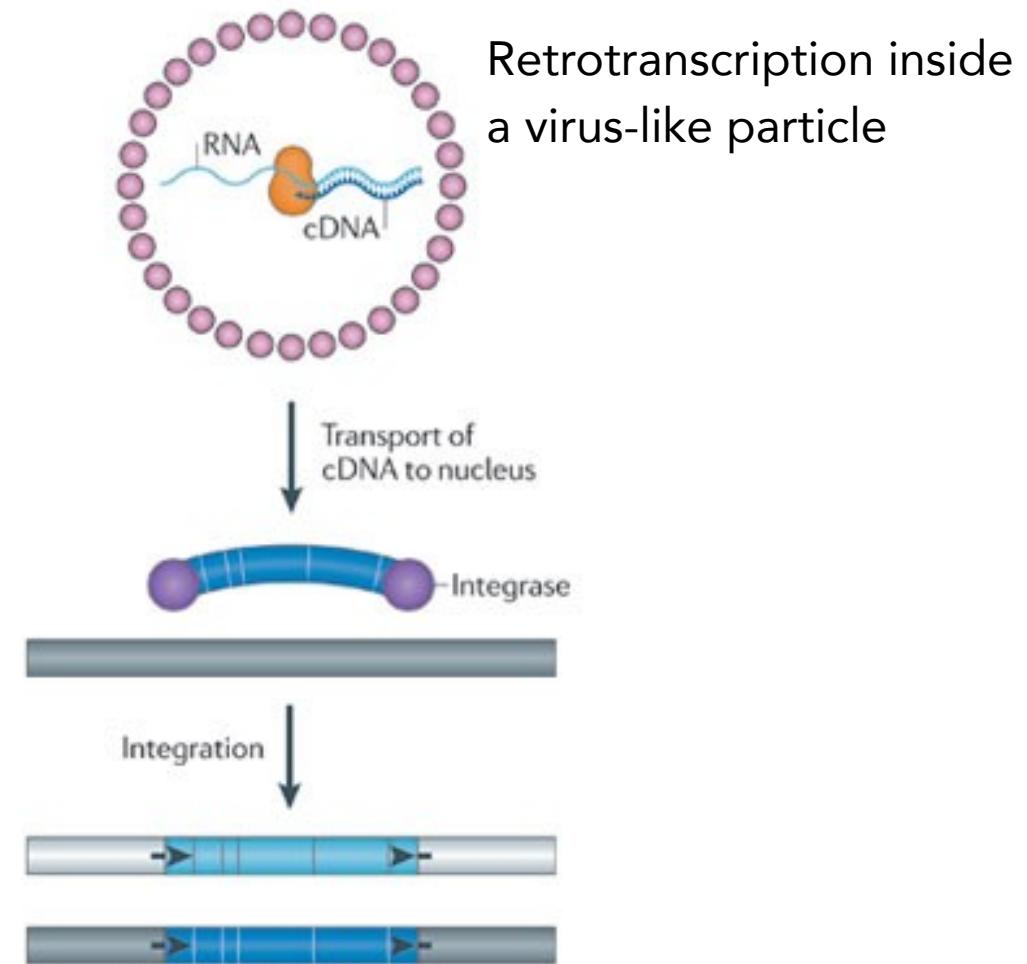
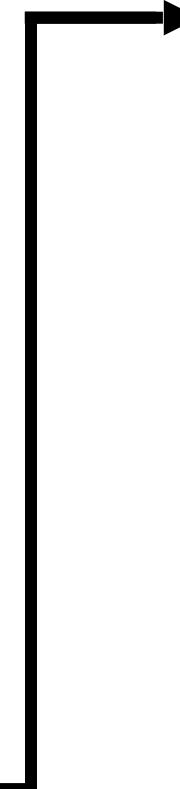
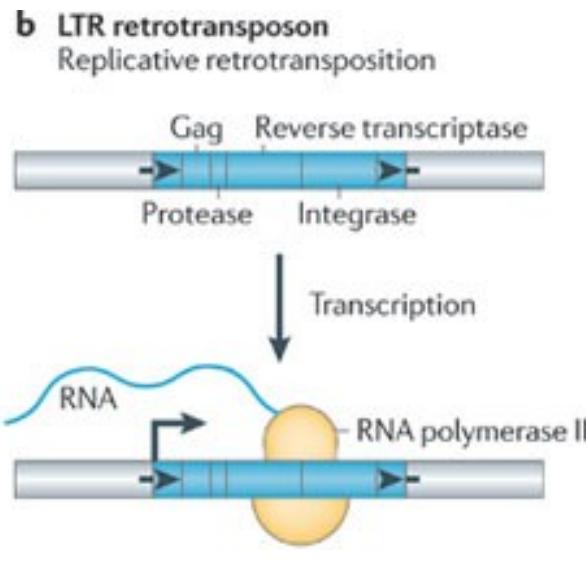
Replicative retrotransposition



Reverse transcription

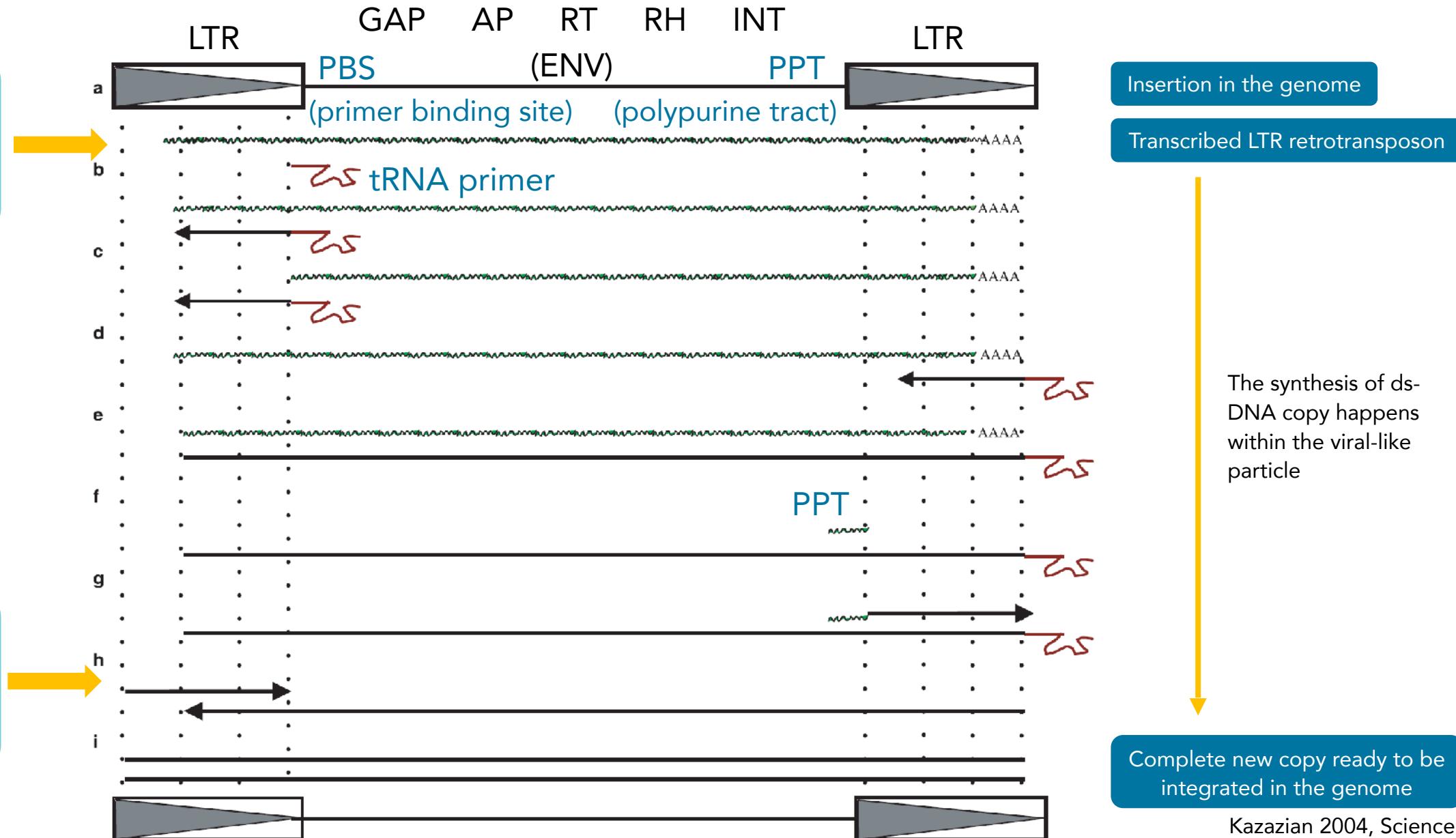


Replicative retrotransposition



LTR are essential for retrotransposition

Transcription starts from an internal promoter!



TE characteristics so far...

- Promoter is necessary → Its integrity is essential!
- Encode proteins and recognition motifs for their own mobilization
- Transpositions leave signatures in the genome: target site duplications
- Their particular mobilization can have different effects on genome evolution:
 - clustering (target site preferentiality)
 - trans-mobilization of other mRNAs (e.g., non-autonomous TEs and retrogenes)
 - trigger of non-allelic homologous (ectopic) recombination

Class II: DNA transposons

Class II (DNA transposons) - Subclass 1

TIR	Tc1-Mariner	Tase*	TA	DTT	P, M, F, O	
	hAT	Tase*	8	DTA	P, M, F, O	
	Mutator	Tase*	9–11	DTM	P, M, F, O	
	Merlin	Tase*	8–9	DTE	M, O	
	Transib	Tase*	5	DTR	M, F	
	P	Tase	8	DTP	P, M	
	PiggyBac	Tase	TTAA	DTB	M, O	
	PIF-Harbinger	Tase*	ORF2	3	DTH	P, M, F, O
	CACTA	Tase	ORF2	2–3	DTC	P, M, F

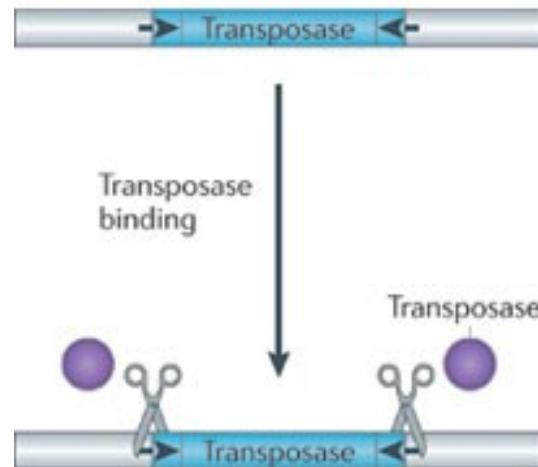
Tase – Transposase

TIR – Terminal Inverted Repeat



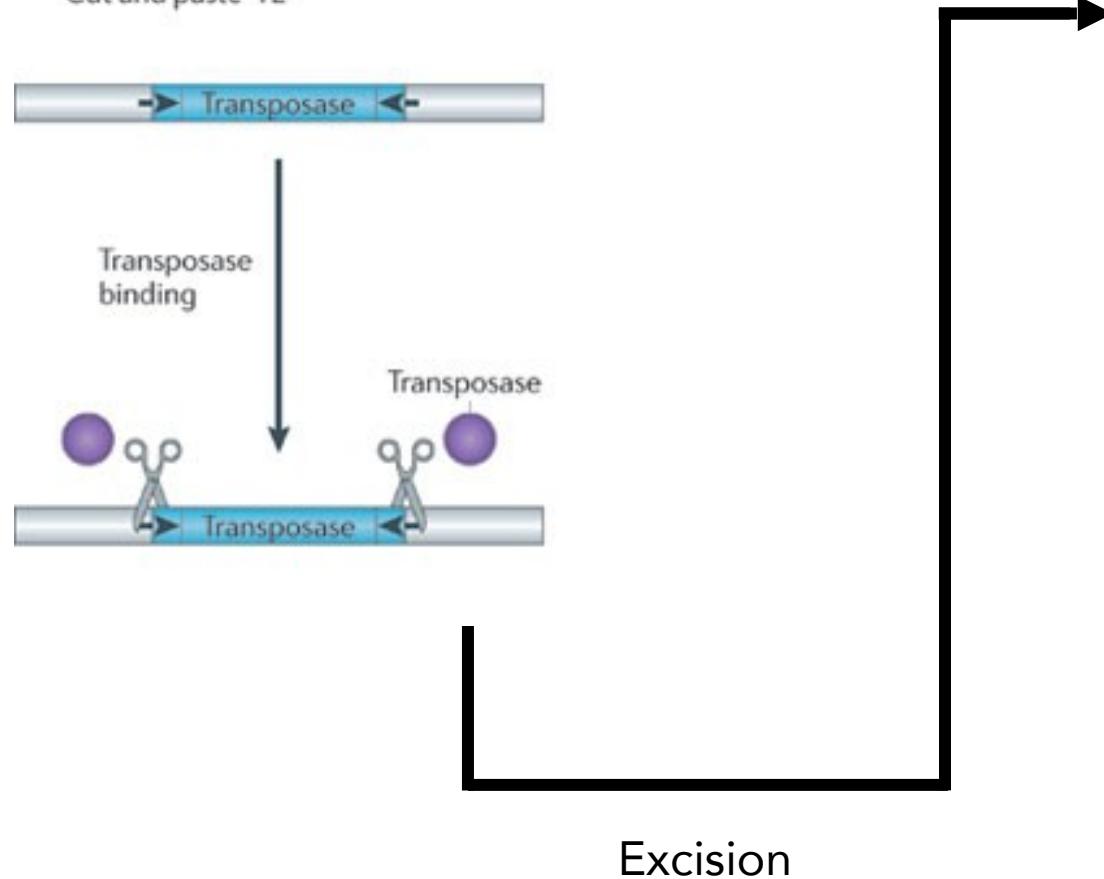
Cut and paste transposition

a DNA transposon
'Cut and paste' TE



Cut and paste transposition

a DNA transposon
'Cut and paste' TE

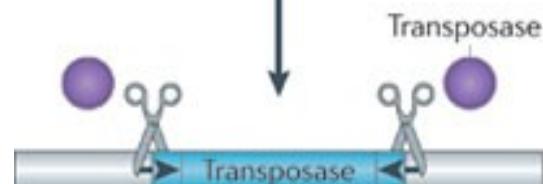


Cut and paste transposition

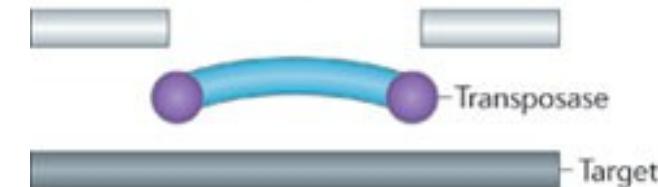
a DNA transposon
'Cut and paste' TE



Transposase binding

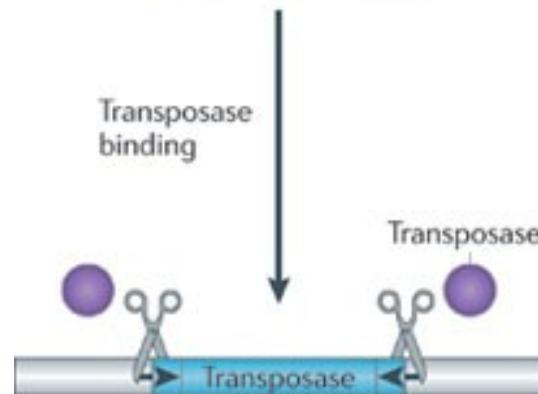


Excision

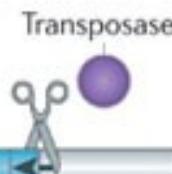


Cut and paste transposition

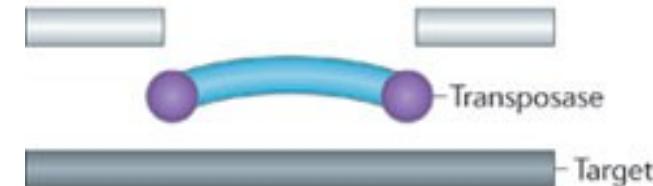
a DNA transposon
'Cut and paste' TE



Transposase
binding



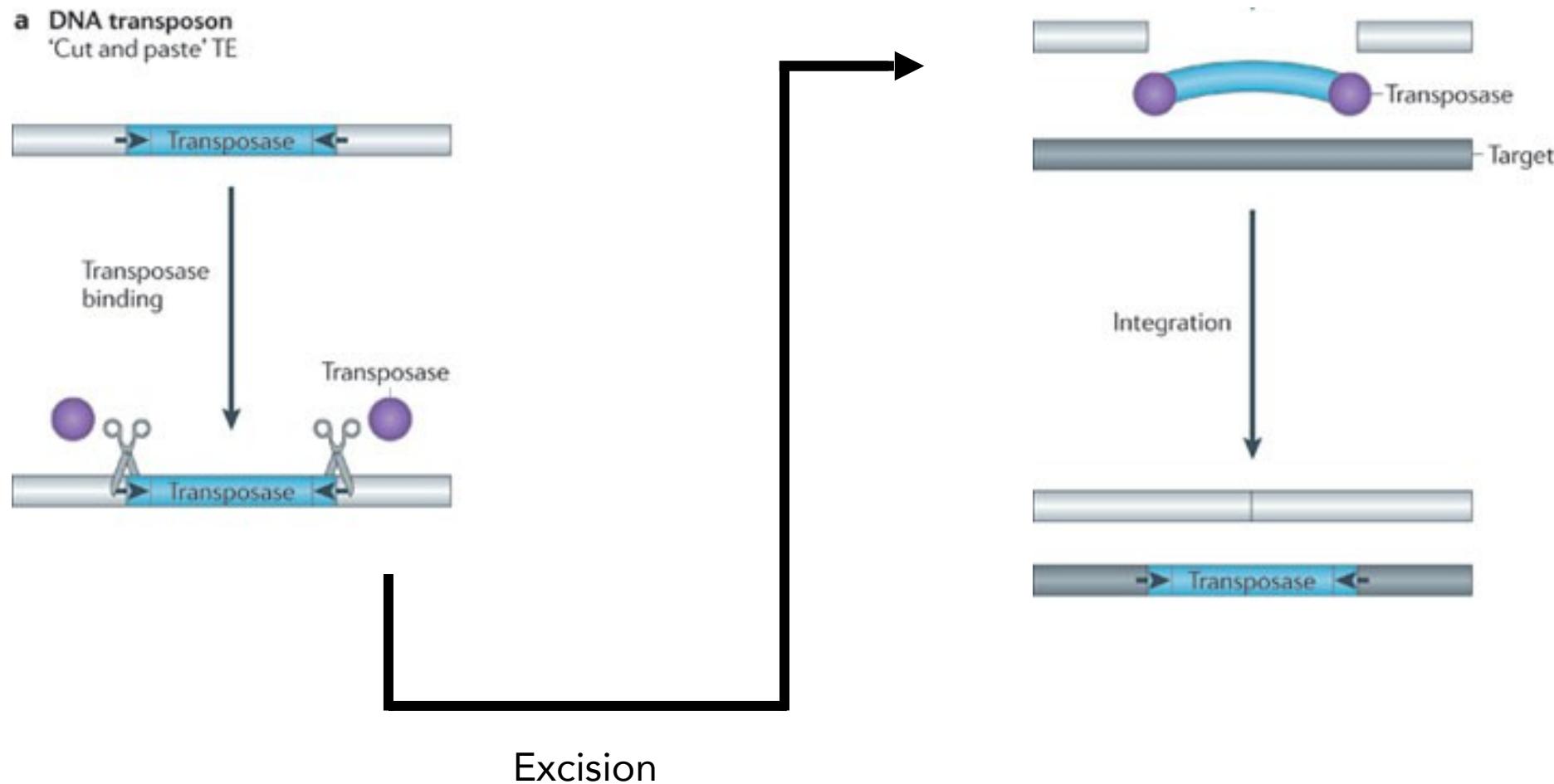
Excision



Integration

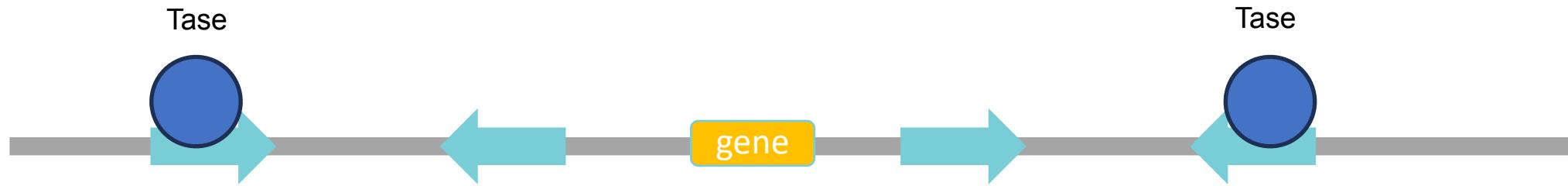


Cut and paste transposition



Mobile DNA

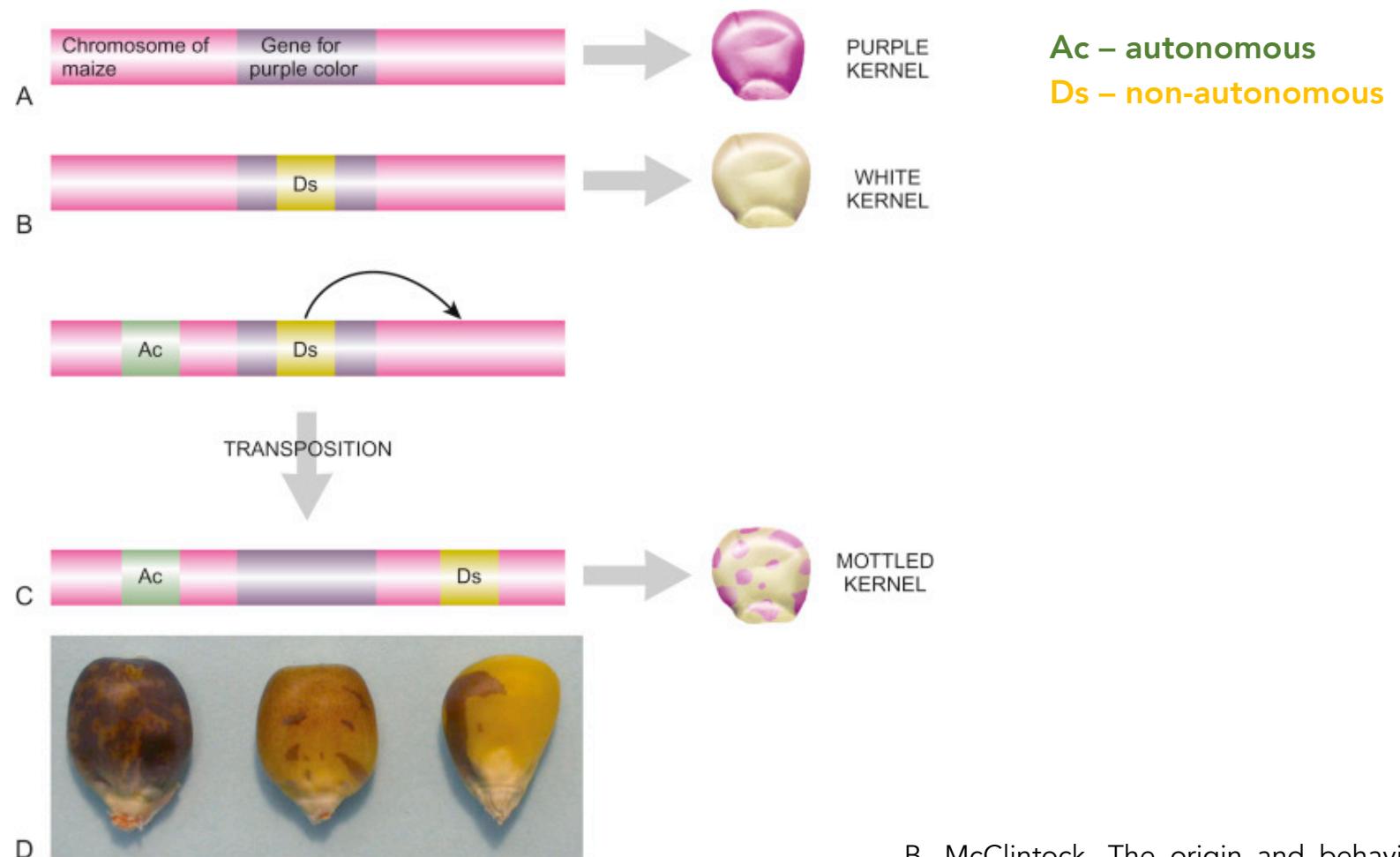
Cut and paste transposition



Translocation of genomic regions and/or genes!

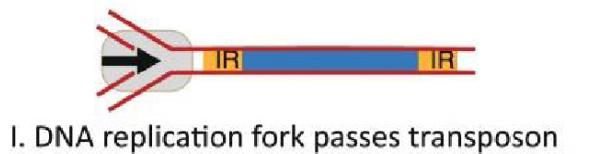
Ac/Ds transposons

hAT family



B. McClintock, The origin and behavior of mutable loci in maize, Proc. Natl. Acad. Sci. U.S.A. 36 (6) 344-355, <https://doi.org/10.1073/pnas.36.6.344>

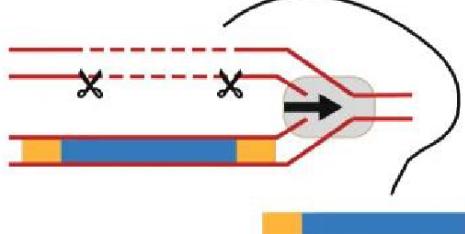
How to increase in copy number?



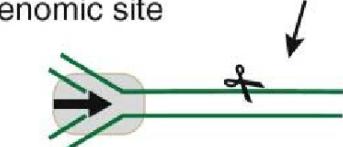
I. DNA replication fork passes transposon



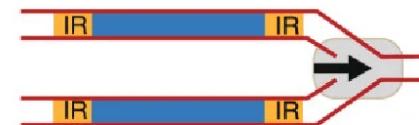
II. Newly replicated transposon is cut out...



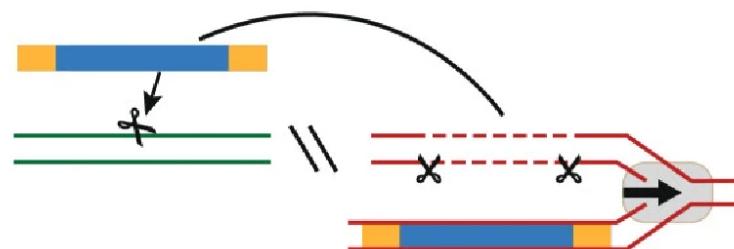
III. ...and inserted into a not-yet replicated genomic site



IV. DNA replication fork passes insertion site



V. Newly replicated transposon is cut out...



VI. ...and transposed into a new locus



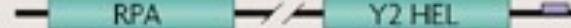
VII. Following transposition, the double-stranded break is repaired by homology-dependent DNA repair

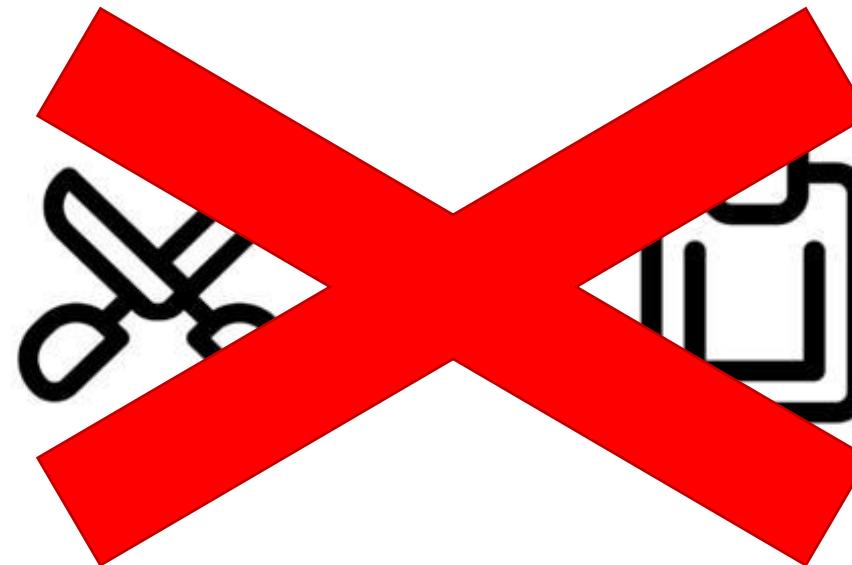


TE characteristics so far...

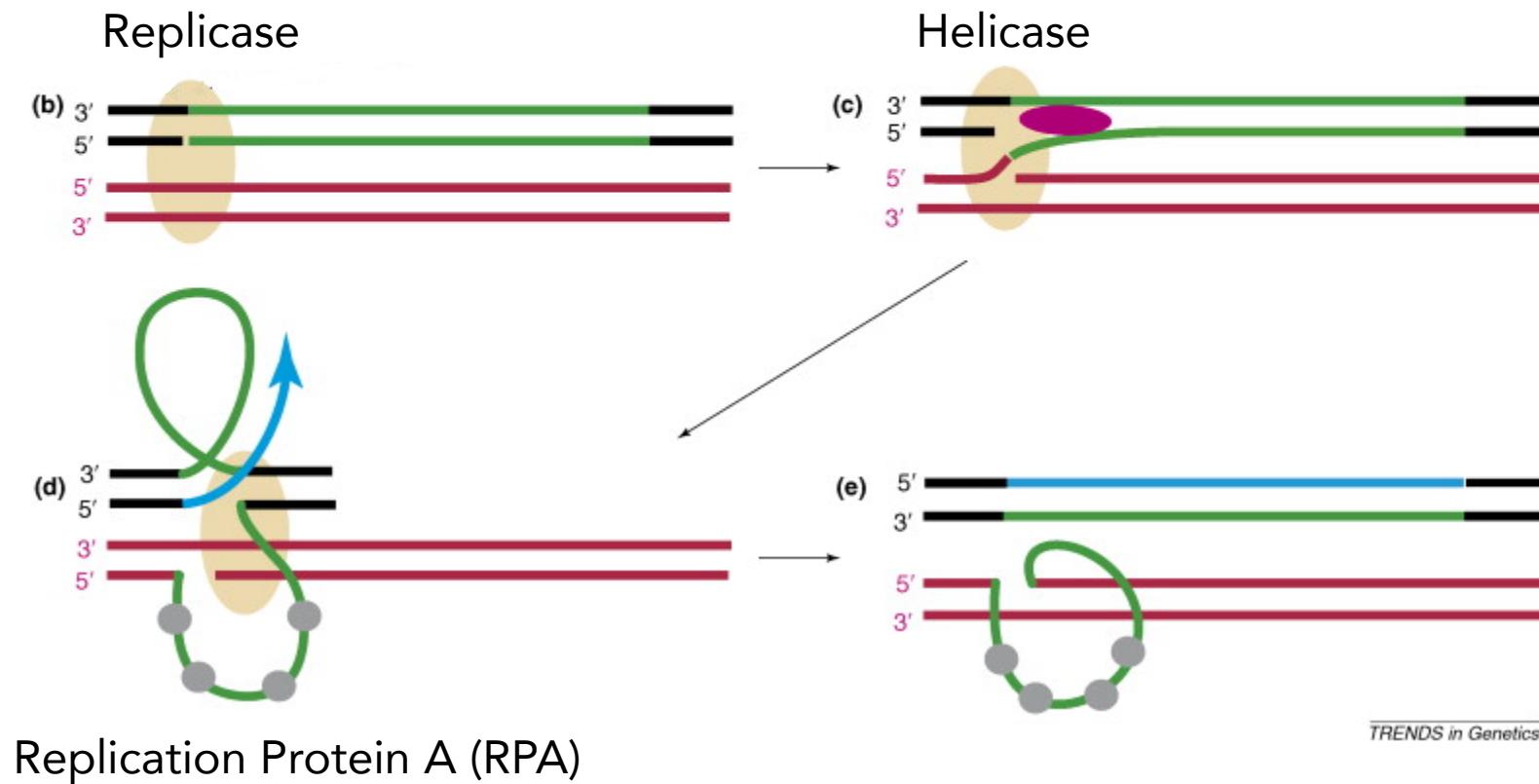
- Promoter is necessary → Its integrity is essential!
- Encode proteins and recognition motifs for their own mobilization
- Transpositions leave signatures in the genome: target site duplications
- Their particular mobilization can have different effects on genome evolution:
 - clustering (target site preferentiality)
 - trans-mobilization of other mRNAs (e.g., non-autonomous TEs and retrogenes)
 - trigger of non-allelic homologous (ectopic) recombination
- Excision-based transposition can also lead to increased copy number

Class II: DNA transposons (subclass 2)

Class II (DNA transposons) - Subclass 2							
Helitron	Helitron		0	DHH	P, M, F		
Maverick	Maverick		6	DMM	M, F, O		

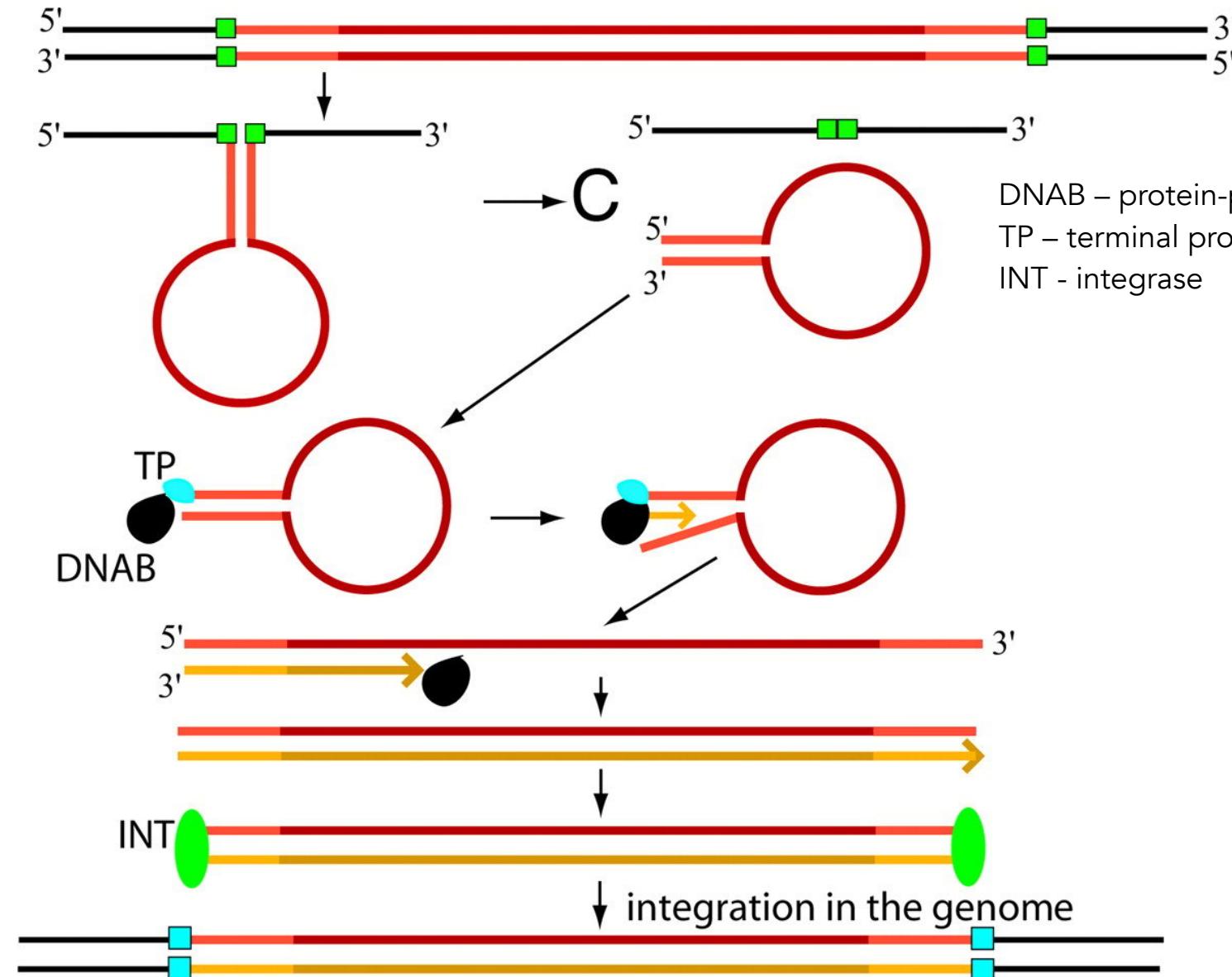


Rolling circle transposition: Helitrons

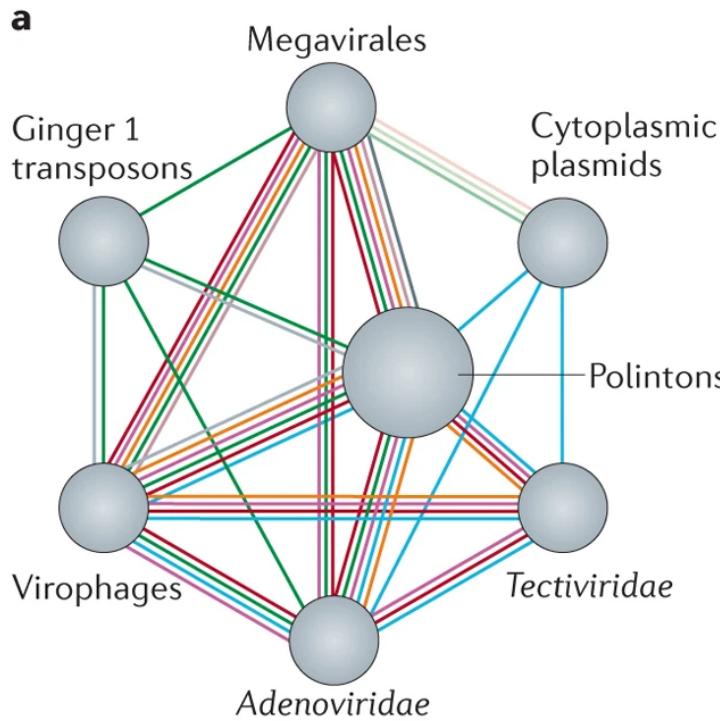


TRENDS in Genetics

Self-synthesizing transposition: Mavericks/Polintons

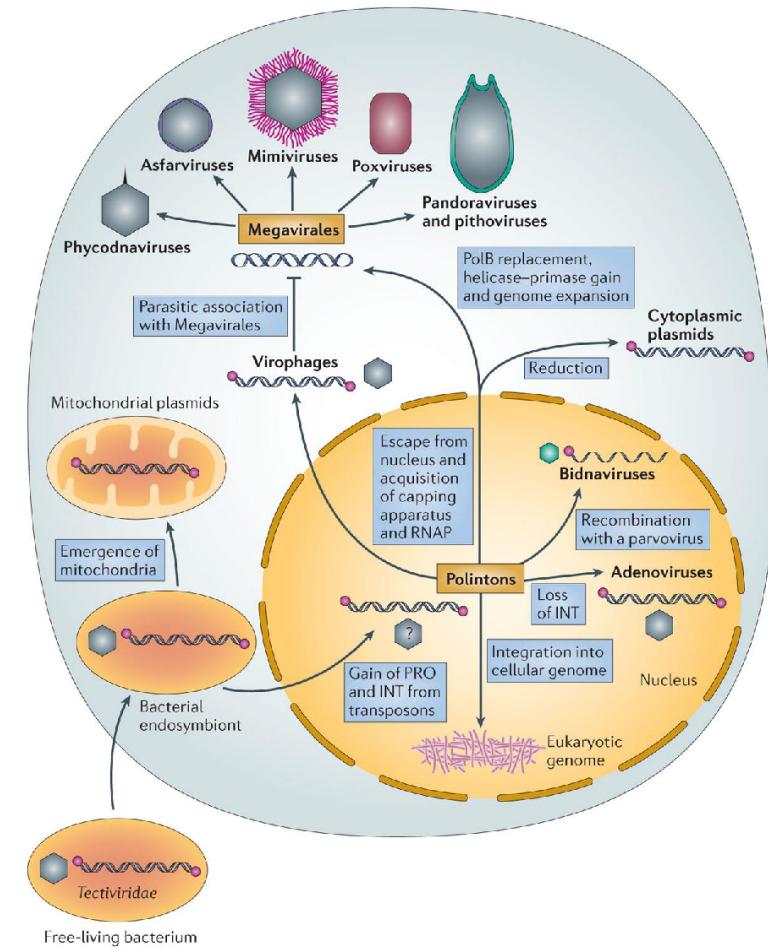


Viral origin of Polintons?

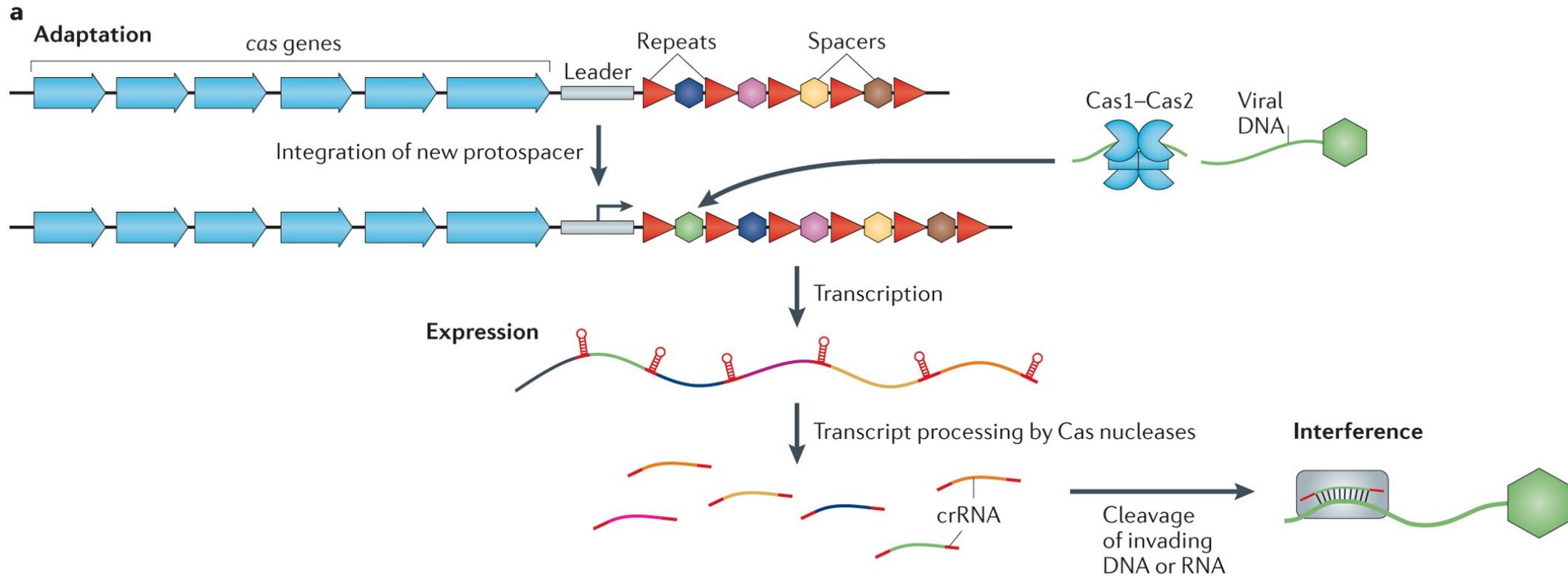


The Polinton genomes are 15–20 kb in size and encode several conserved proteins, including:

- protein-primed type B DNA polymerase (pPolB)
- retroviral-like (RVE) family integrase
- FtsK-like ATPase
- adenovirus-type cysteine protease
- two putative capsid proteins

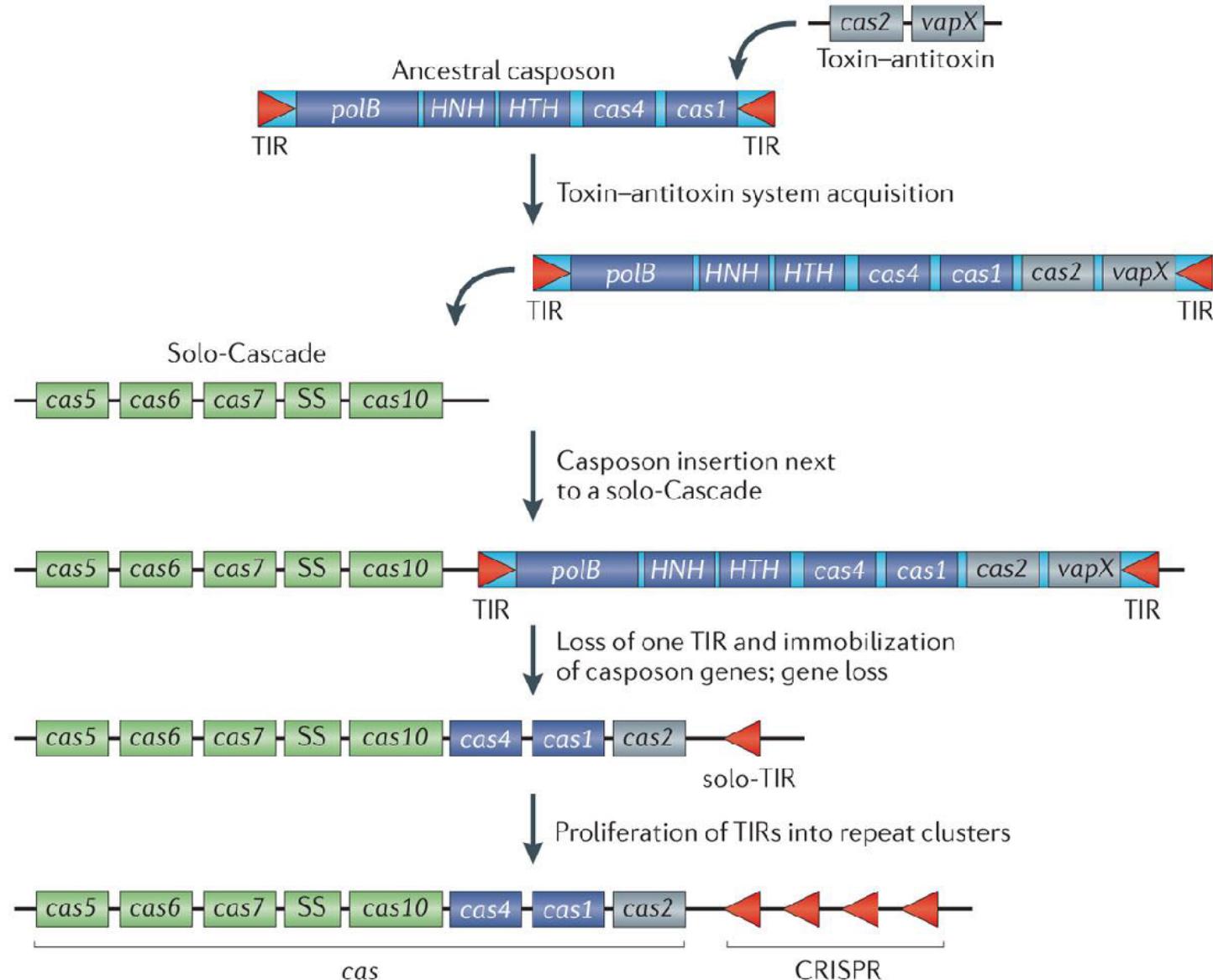


CRISPR-Cas and transposons



a The prokaryotic clustered regularly interspaced short palindromic repeat–CRISPR-associated protein (CRISPR–Cas) locus consists of cas genes (blue arrows) that encode different Cas proteins, and CRISPR arrays composed of variable spacers (coloured hexagons) interspersed with direct repeats (red triangles). The leader sequence (grey rectangle) contains a promoter for the transcription of the CRISPR array and marks the end where new spacers are incorporated. Three stages of CRISPR–Cas immunity are depicted. During the adaptation stage, a Cas1–Cas2 heterohexamer uptakes a protospacer from the invading plasmid or viral DNA (green) and incorporates it at the leader-proximal end of the CRISPR array. During the expression stage, the CRISPR array is transcribed, and the transcript is processed into small CRISPR RNAs (crRNAs) by different Cas nucleases in a CRISPR–Cas type-dependent manner. During the interference stage, crRNAs act as guides for the cleavage of invading viral or plasmid DNA or RNA that contains regions complementary to the crRNA

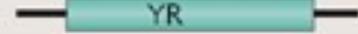
CRISPR-Cas and transposons



TE characteristics so far...

- Promoter is necessary → Its integrity is essential!
- Encode proteins and recognition motifs for their own mobilization
- Transpositions leave signatures in the genome: target site duplications
- Their particular mobilization can have different effects on genome evolution:
 - clustering (target site preferentiality)
 - trans-mobilization of other mRNAs (e.g., non-autonomous TEs and retrogenes)
 - trigger of non-allelic homologous (ectopic) recombination
- Excision-based transposition can also lead to increased copy number
- TEs can originate from viruses and viceversa (e.g., LTR retrotransposons and Polintons)

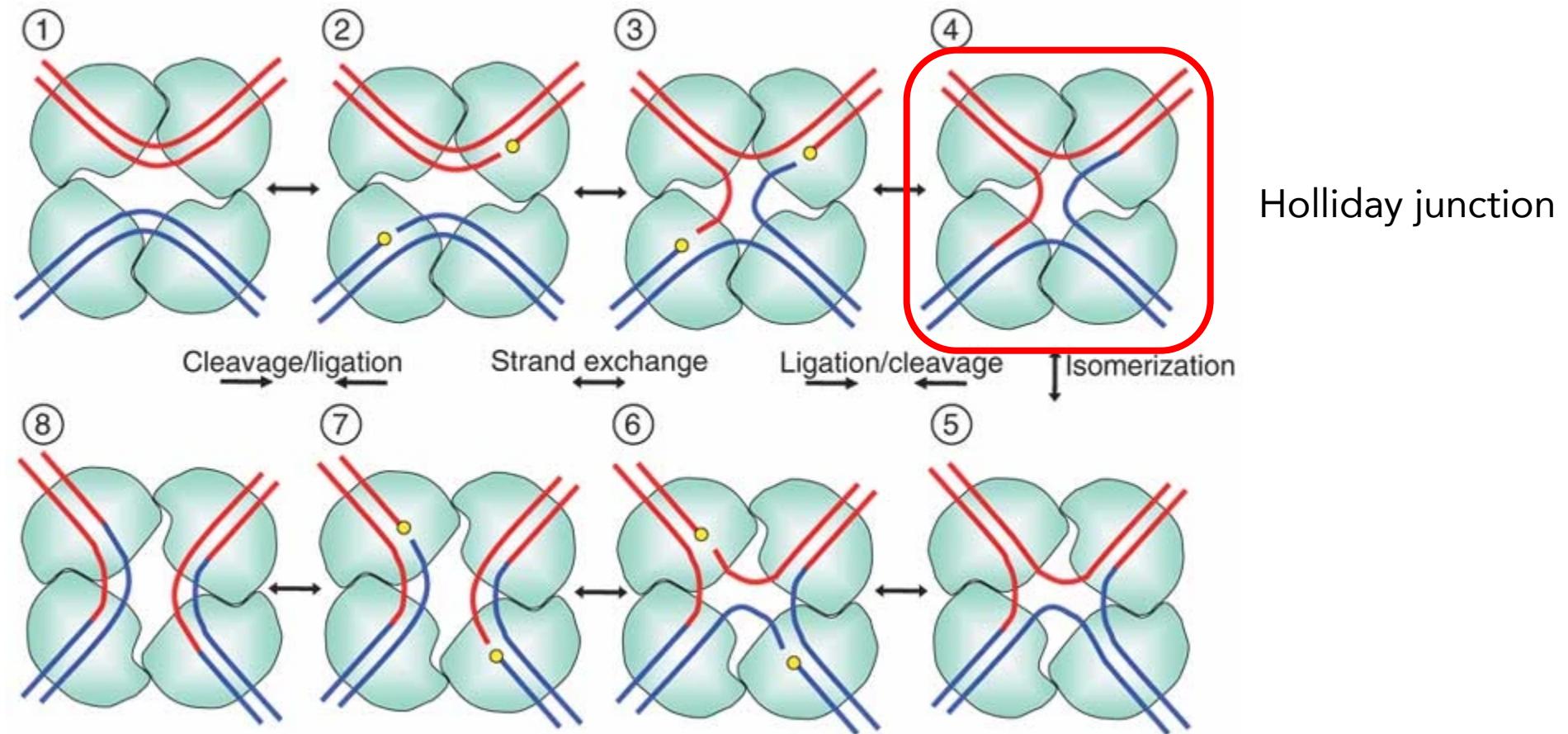
TEs with Tyrosine Recombinase

Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
Class I (retrotransposons)					
DIRS	DIRS		0	RYD	P,M,F,O
Class II (DNA transposons) - Subclass 2					
Crypton	Crypton		0	DYC	F

TE integration mechanism occurs via:

- Endonuclease: [LINE](#), [SINE](#), [PLE](#)
 - DDE-Transposase: [TIR](#)
 - Integrase: [LTR](#), [Maverick/Polinton](#)
 - Rep protein: [Helitron](#)
 - Tyrosine recombinase: [DIRS](#), [Crypton](#)
- [Class I: retrotransposons](#)
[Class II: DNA transposons](#)

Site-specific recombination



Holliday junction

DIRS and Cryptons have no TSDs (with exceptions)

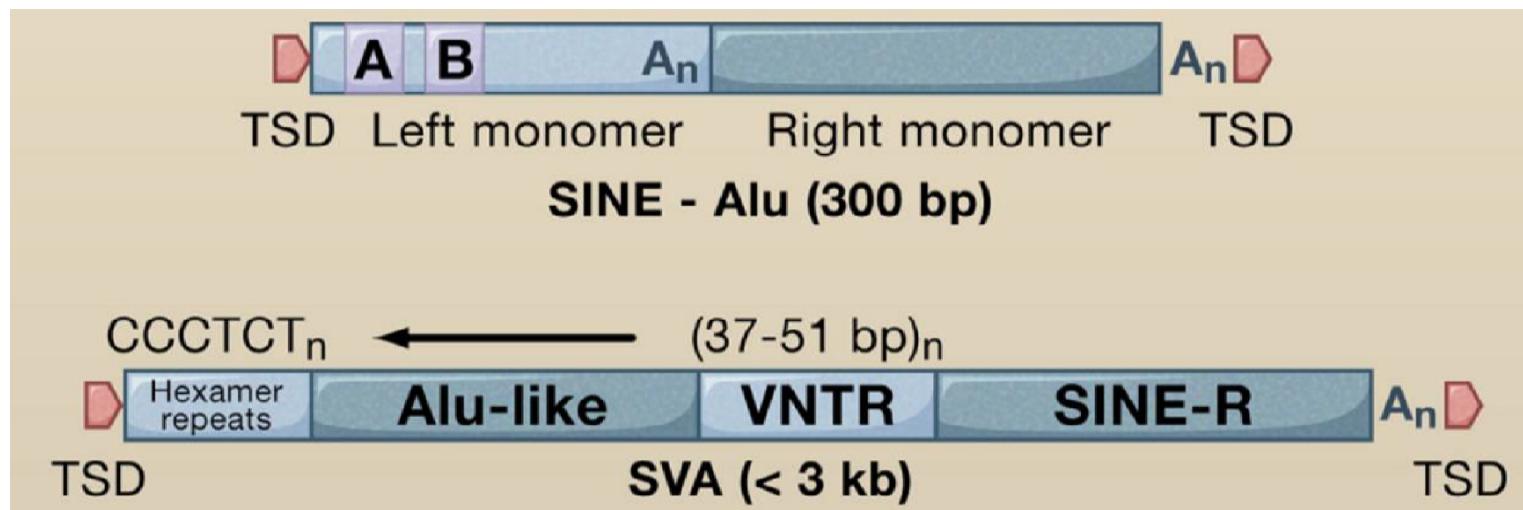
Rice 2005, Nat Struct Mol Bio

TE characteristics so far...

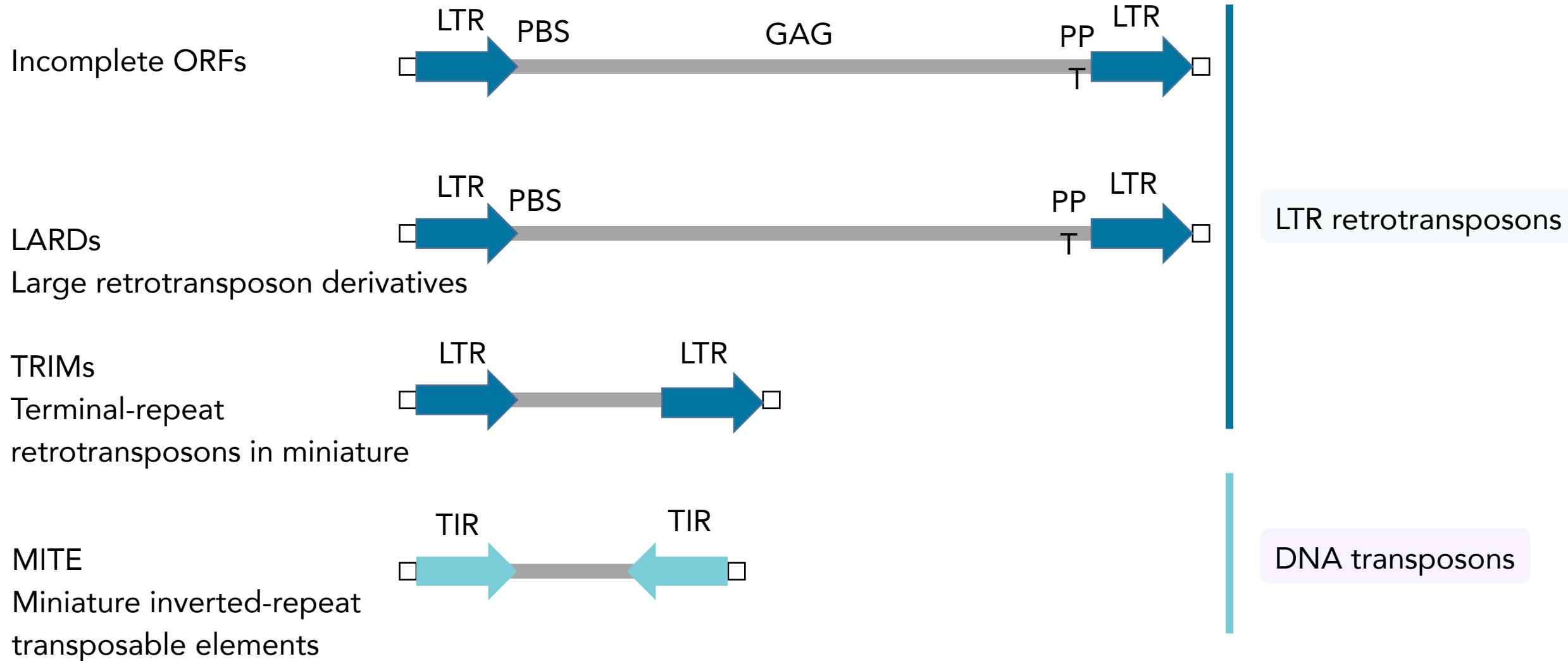
- Promoter is necessary → Backup needed, e.g., LTRs
- Encode proteins and recognition motifs for their own mobilization
- Transpositions leave signatures in the genome: target site duplications
- Their particular mobilization can have different effects on genome evolution:
 - clustering (target site preferentiality)
 - trans-mobilization of other mRNAs (e.g., non-autonomous TEs and retrogenes)
 - trigger of non-allelic homologous (ectopic) recombination
- Excision-based transposition can also lead to increased copy number
- TEs can originate from viruses and viceversa (e.g., LTR retrotransposons and Polintons)
- Human classification of TEs not always can account for complex evolutionary origins

Non-autonomous TEs

Non-LTR retrotransposons

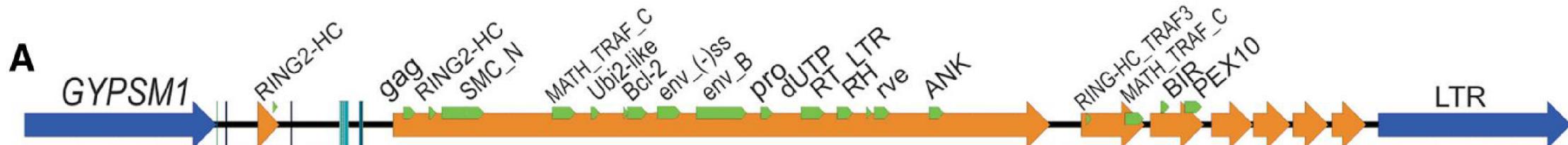


Anything can be a non-autonomous TEs



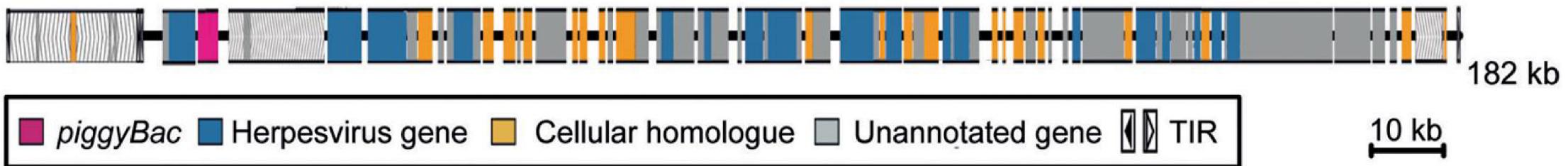
Giant TEs

LTR retrotransposons: *Burro*(30 kb, planaria *Schmidtea mediterranea*)



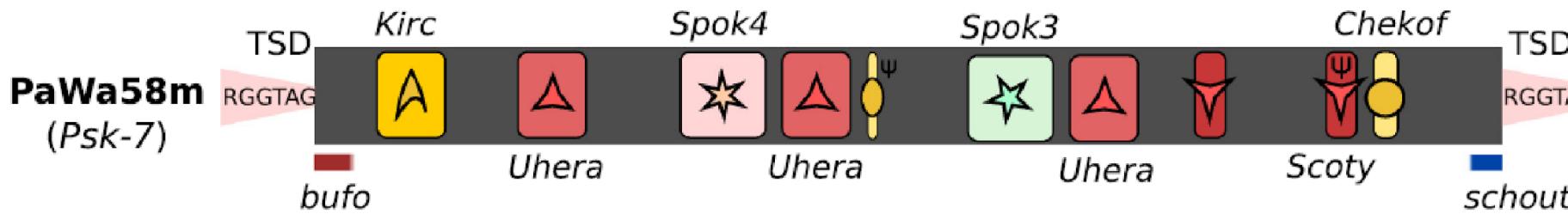
Grohme et al 2018, Nature

TIR DNA transposons: *Teratorn*(182 kb, medaka *Oryzias latipes*)



Arkhipova 2019, Genome Biol. Evol.

YR DNA transposons: *Enterprise*(247 kb, fungus *Podospora anserina*)



Vogan et al. 2021, Genome Res.

Mobility mechanisms define the upper size limit of TEs

TEs are selfish elements

Selfish genetic elements

(anything ranging from single genes or chromosomes to entire genomes)

=

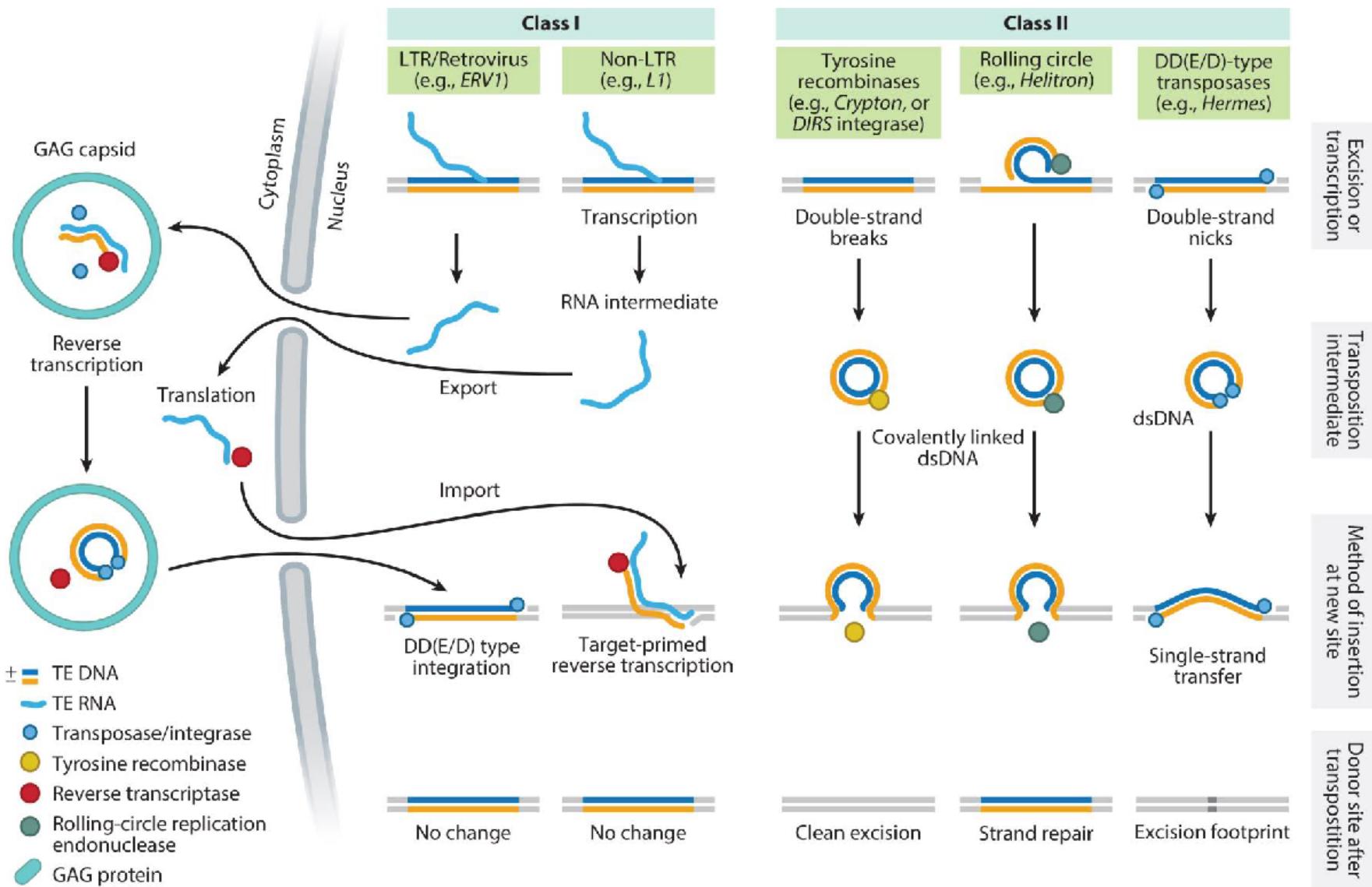
Genetic element with the sole “ purpose” to transmit itself

(which often comes with a cost to its host)
“When a given DNA, or class of DNAs, of unproven phenotypic function can be shown to have evolved a strategy (such as transposition) which ensures its genomic survival, then no other explanation for its existence is necessary. The search for other explanations may prove, if not intellectually sterile, ultimately futile.” According to this theory, the evolutionary success of transposable elements is not attributed to a specific cellular function, as McClintock envisioned, but simply to their ability to replicate independently of the rest of the genome, thereby ensuring their propagation without benefitting the organism, but sometimes at its expense — much like viruses.

Feschotte, C. Transposable elements:
McClintock’s legacy revisited. Nat Rev
Genet 24, 797–800 (2023). <https://doi.org/10.1038/s41576-023-00652-3>

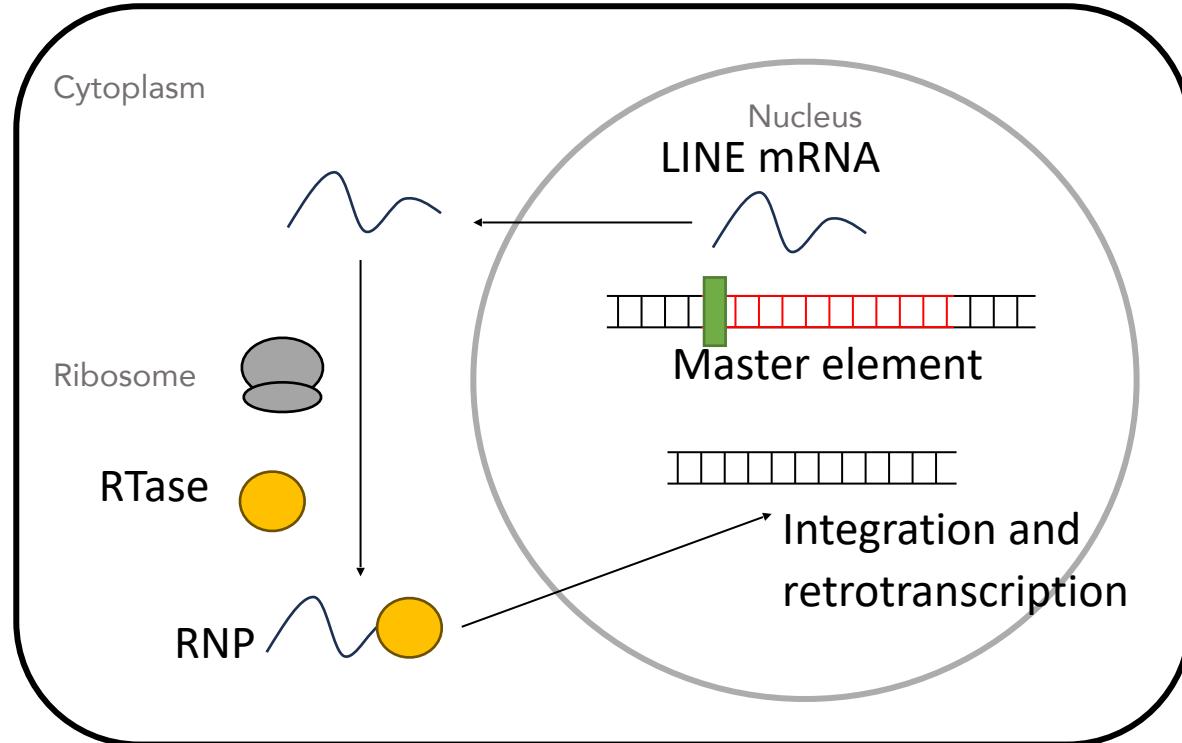
Doolittle, W. F. & Sapienza, C. Selfish
genes, the phenotype paradigm and
genome evolution. Nature 284, 601–603
(1980).

Mechanisms overview



LINE retrotransposons

Where/when/how in the cell



Target site preference and TSD

Target site preferentiality for sequences similar to the 3' UTR

Target site duplications are of variable length

Requirements for mobility

Transcription: promoter for pol II -> mRNA + polyA

Replication: RTase

Recognition site for cis-mobilisation

Integration: endonuclease

No introns

Content of new copies

Mother copy



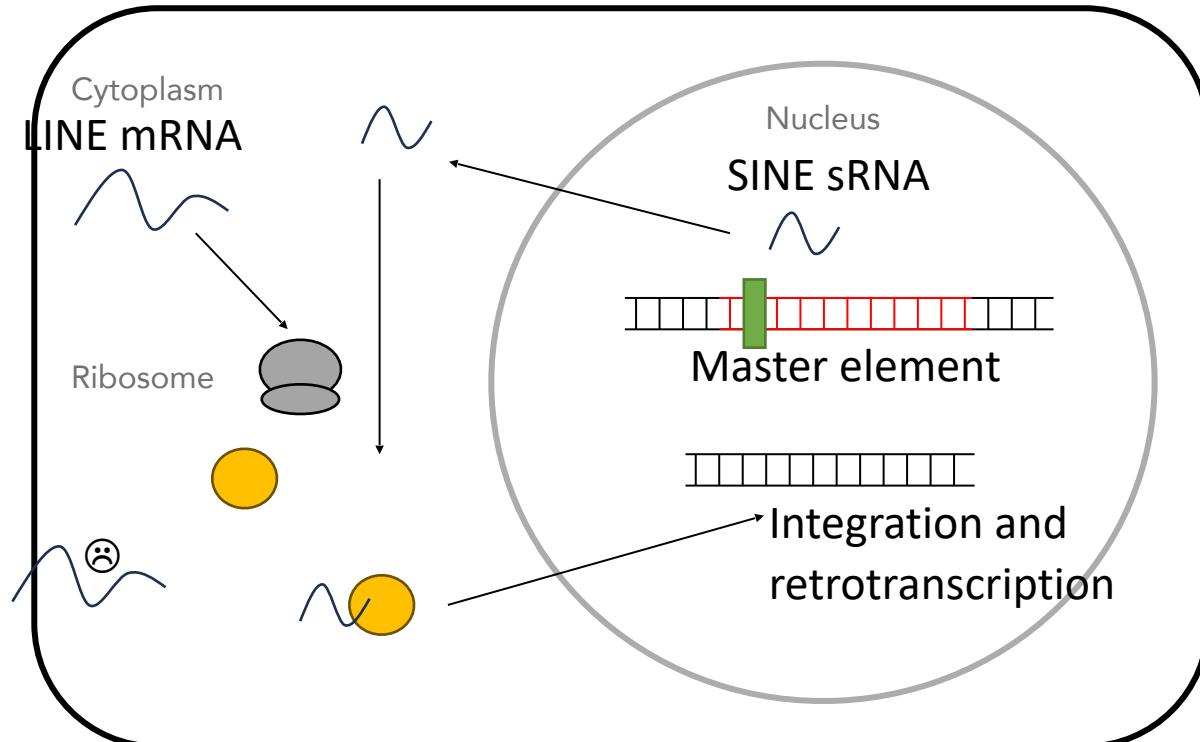
Daughter copies



5' truncation
“dead on arrival”

SINE retrotransposons

Where/when/how in the cell



Target site preference and TSD

Target site preferentiality for sequences similar to the 3' UTR

Target site duplications are of variable length

Requirements for mobility

Transcription: promoter for pol III → sRNA

Replication + integration: using LINE derived proteins

Recognition site for trans-mobilisation

No introns

Content of new copies

Mother copy



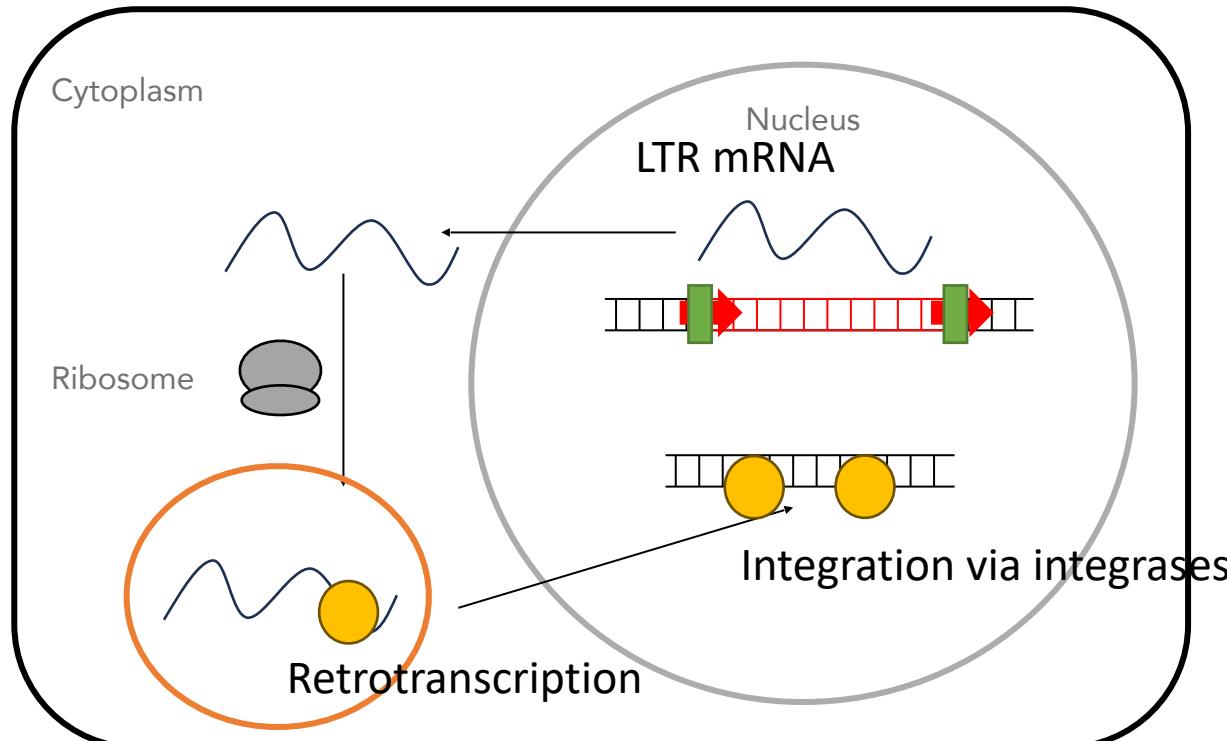
Daughter copies



5' truncation
“dead on arrival”

LTR retrotransposons

Where/when/how in the cell



Target site preference and TSD

No preferentiality for target site

Specific length of target site duplications (4 bp, 5 or 6 bp)

Requirements for mobility

Transcription: promoter for pol II -> mRNA + polyA

Replication: within viral-like particle, gag (capsid), protease, RTase, RNase H

Integration: integrase

Recognition site for cis-mobilisation (LTR)

No introns

Content of new copies

Mother copy



Daughter

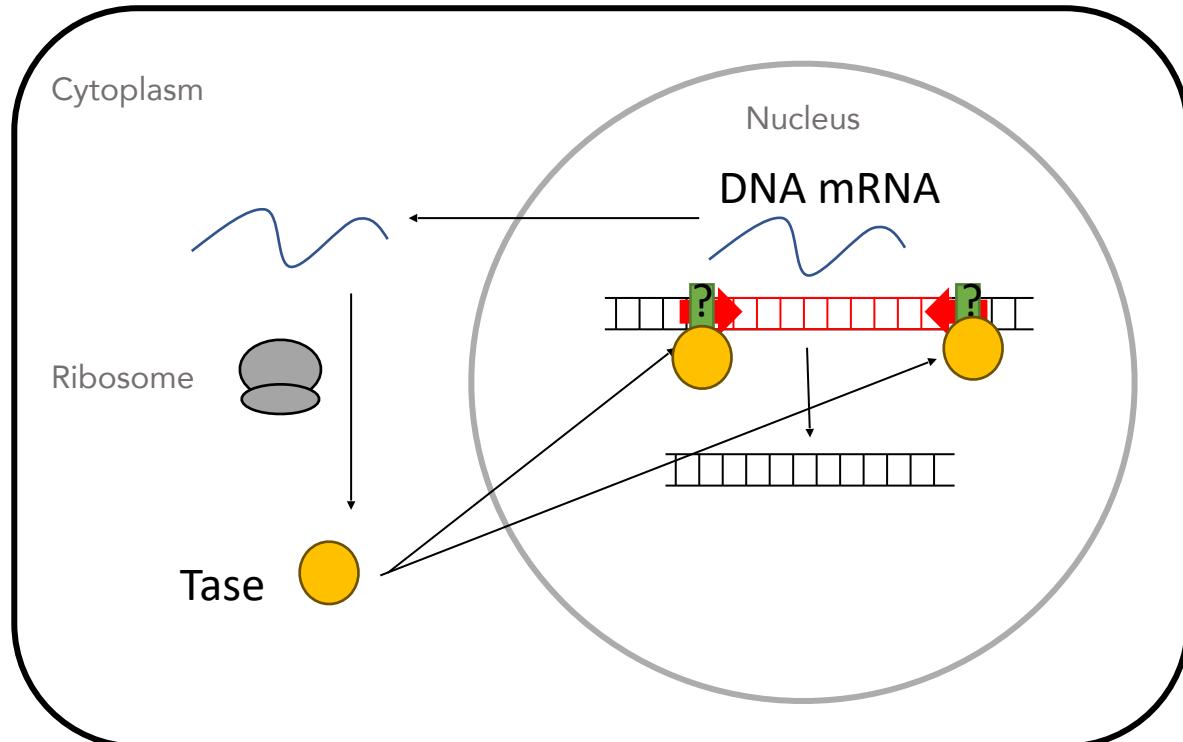


NAHR

Solo-LTR

DNA transposons (TIR)

Where/when/how in the cell



Target site preference and TSD

There can be specificity for target site (e.g., TA, TTAA) and there can be specific target site duplication length (e.g., 8 bp)

Requirements for mobility

Transcription: promoter for pol II → mRNA + polyA

Mobilisation and integration: transposase (Tase)

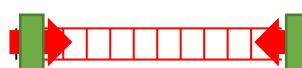
Replication: dependent on host DNA replication

Recognition site in TIRs allows for both cis- and trans-mobilisation (with different probabilities)

Might have introns

Content of new copies

Mother copy



Daughter copy



Inclusion of extra DNA
or loss of ORFs

Group discussion

1. How do LINE/SINE/LTR/DNA (retro)transposons get transcribed, replicated, and integrated?
2. Where do these TEs jump, what happens to the new insertion site, and what happens to the new insertion?
3. What does every new insertion contain or not contain and what consequences does this have for TE and host?
4. In what state do these TEs show up in the nucleus or cytoplasm and in the transcriptome or proteome?