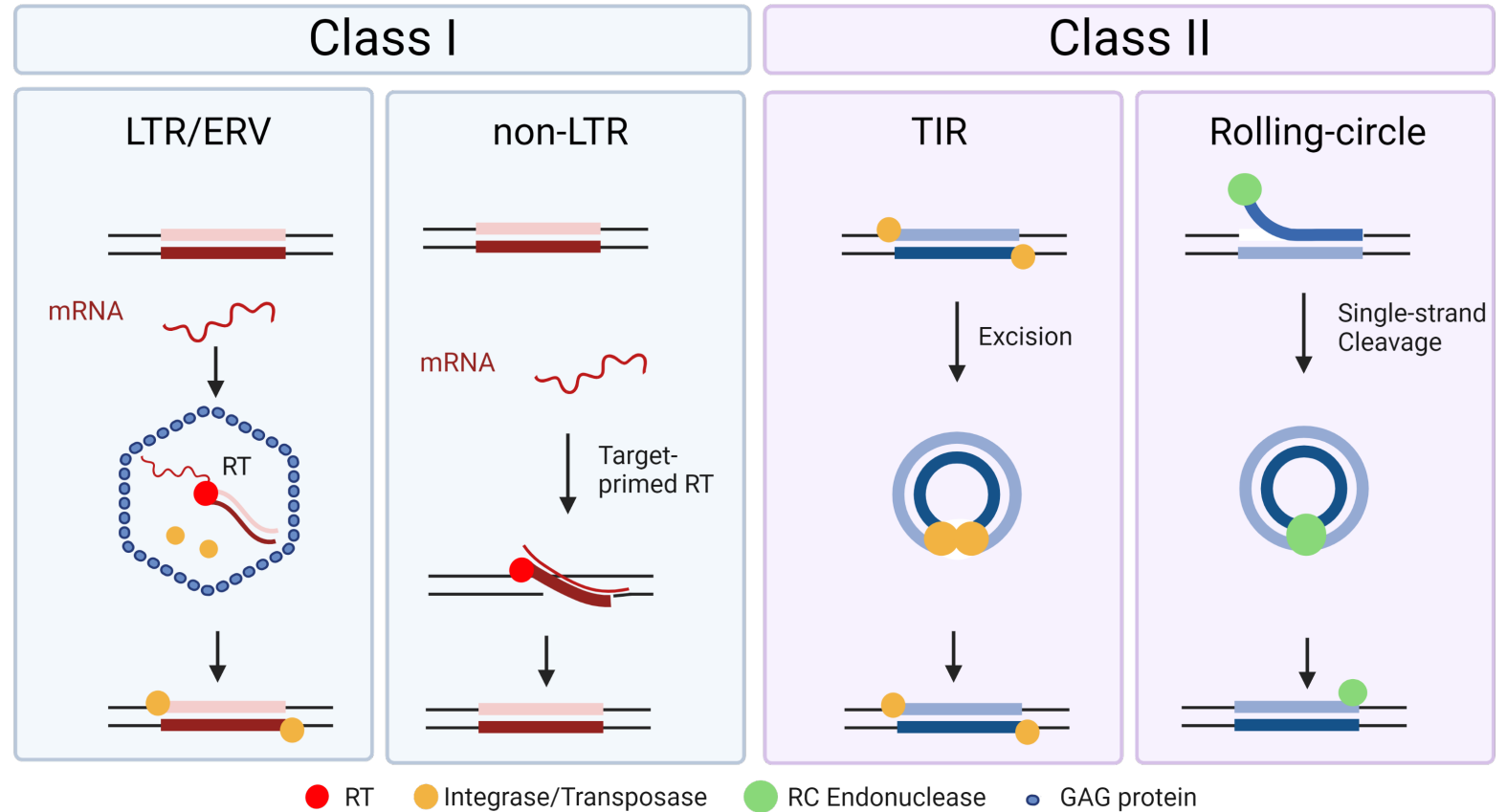
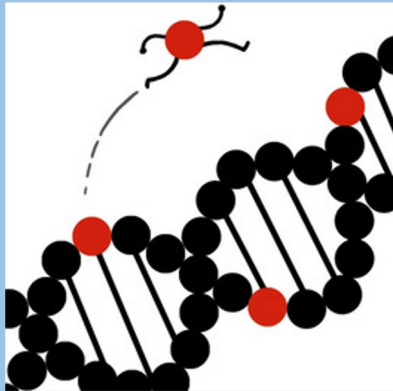


TE diversity recap

A FIELD GUIDE TO THE TEs OF THE WORLD



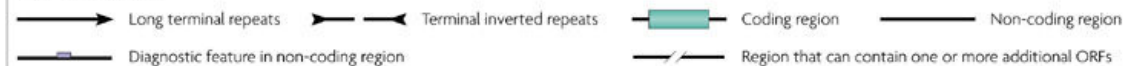
How do LINE/SINE/LTR/DNA (retro)transposons get transcribed, replicated, and integrated?

Where do these TEs jump, what happens to the new insertion site, and what happens to the new insertion?

FEATURE	LINE	SINE	LTR	DNA (TIR)
Class	Class I	Class I	Class I	Class II
Structural features (e.g., internal repeats, ORFs)	one or multiple ORFs; poly-A tails or microsatellite tail	lack ORFs; promoter pol III; recognition motif at 3' end	promoter in the LTRs; Long Terminal Repeats usually start with 5' –TG and end with CA-3' ; direct repeats	Inverted terminal repeats; 1 transposase ORF
Transcription	promoter pol II	promoter pol III (tRNA, ribosomal promoter)	promoter pol II	promoter pol II
Replication mechanism	Target-primed reverse transcription (TPRT); tail is used as a primer for RT	TPRT	Retrotransposition with viral particle (RT)	Non-replicative transposition
Integration enzyme	Endonucleases	Endonuclease of LINEs	Integrase	Transposase
Mobilization in cis/trans	cis	trans	cis	trans (very unlikely mobilises the original element in cis)
Increase in copy number	part of the transposition mechanism	part of the transposition mechanism	part of the transposition mechanism	increase in copy number dependent on the host cell cycle
Presence of introns	no	no	no	it could/should
Target site preferentiality	depends on the tail sequence – sometimes can be very specific (e.g., R2) – usually not very specific	depends on the tail	no preferentiality	there can be preferentiality: target site motif
Target site duplication (upon transposition)	yes; variable length	yes; variable length	precise length: 4, 5, 6 nt TSD	variable length; precise length; motif (e.g., "TA", "TTAA")
State of the new insertion	5' truncation; daughter copies can be truncated; "dead on arrival"	5' truncation; daughter copies can be truncated; "dead on arrival"	copy of the mother element; they can full-length or solo-LTRs	they are the original element mother = daughter copy; transpose nearby genomic regions; new copy has TSDs and the old locus is left with the TSDs

Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
Class I (retrotransposons)					
LTR	Copia	→ GAG AP INT RT RH →	4-6	RLC	P, M, F, O
	Gypsy	→ GAG AP RT RH INT →	4-6	RLG	P, M, F, O
	Bel-Pao	→ GAG AP RT RH INT →	4-6	RLB	M
	Retrovirus	→ GAG AP RT RH INT ENV →	4-6	RLR	M
	ERV	→ GAG AP RT RH INT ENV →	4-6	RLE	M
DIRS	DIRS	→ GAG AP RT RH YR ←	0	RYD	P, M, F, O
	Ngaro	→ GAG AP RT RH YR → >	0	RYN	M, F
	VIPER	→ GAG AP RT RH YR → >	0	RYV	O
PLE	Penelope	← RT EN →	Variable	RPP	P, M, F, O
LINE	R2	RT EN	Variable	RIR	M
	RTE	APE RT	Variable	RIT	M
	Jockey	ORF1 APE RT	Variable	RIJ	M
	L1	ORF1 APE RT	Variable	RIL	P, M, F, O
	I	ORF1 APE RT RH	Variable	RII	P, M, F
SINE	tRNA		Variable	RST	P, M, F
	7SL		Variable	RSL	P, M, F
	5S		Variable	RSS	M, O
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	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 proteinase APE, Apurinic endonuclease ATP, Packaging ATPase C-INT, C-integrase CYP, Cysteine protease EN, Endonuclease
 ENV, Envelope protein GAG, Capsid protein HEL, Helicase INT, Integrase ORF, Open reading frame of unknown function
 POL B, DNA polymerase B RH, RNase H RPA, Replication protein A (found only in plants) RT, Reverse transcriptase
 Tase, Transposase (* with DDE motif) YR, Tyrosine recombinase Y2, YR with YY motif

Species groups

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

Table 1 Classification and characteristics of eukaryotic DNA transposons

Superfamily	Related IS	TSD	Length ¹ (kb)	TIRs ¹ (bp)	Terminal motif (5'-3')	TPase ¹ (aa)	Catalytic motif	DNA-binding motif	Additional proteins
Tc1/mariner	IS630	TA	1.2–5.0	17–1100	Variable	300–550	DD(30–41)D/E	HTH (cro/paired)	
<i>bAT</i>	nd	8 bp	2.5–5	10–25	YARNG	600–850	D(68)D(324)E ²	ZnF (BED)	
P element	nd	7/8 bp	3–11	13–150	CANRG	800–900	D(83)D(2)E(13)D ³	ZnF (THAP)	
<i>MuDR/ Foldback</i>	IS256	7–10 bp	1.3–7.4	0-sev. Kb	Variable	450–850	DD(~110)E	ZnF (WRKY/GCM1)	
<i>CACTA</i>	nd	2/3 bp	4.5–15	10–54	CMCWR	500–1200	nd	nd	TNPA (DNA- binding protein)
<i>PiggyBac</i>	IS1380	TTAA	2.3–6.3	12–19	CCYT	550–700	DDE?	nd	
<i>PIF/ Harbinger</i>	IS5	TWA	2.3–5.5	15–270	GC-rich	350–550	DD(35–37/ 47–48)E	HTH	PIF2p (Myb/SANT domain)
<i>Merlin</i>	IS1016	8/9 bp	1.4–3.5	21–462	GGNRM	270–330	DD(36–38)E	nd	
<i>Transib</i>	nd	5 bp	3–4	9–60	CACWATG	650–700	DD(206–214)E	nd	
<i>Banshee</i>	IS481	4/15 bp	3–5	41–950	TGT	300–400 ⁴	DD(34)E	HTH	
<i>Helitron</i>	IS91	none	5.5–17	none	5'- TC...CTAR- 3'	1400–3000 ⁵	HHYY (“REP motif”)	ZnF-like	RPA (in plants)
<i>Maverick</i>	none	5/6 bp	15–25	150–700	Simple repeat	350–450 ⁴	DD(33–35)E	ZnF (HHCC)	4–10 DNA virus-like proteins

¹Refers to a potentially complete, autonomous element.²Motif in *Hermes* TPase.³Motif in *Drosophila* P element TPase.⁴RVE integrase-like.⁵REP-Helicase.

nd = not determined.