TRANSPOSONS IN FRUIT FLY

Guillermo Lendrino Regalón

1. INTRODUCTION

Transposons or transposable elements are mobile genetic elements that move, or transpose, to different locations throughout the genomes in which they reside.

These "jumping genes" impact the host genome size by transposition and deletion events, but may also adopt unique functional and evolutionary roles. There are 2 main classes of transposons: DNA transposons and retrotransposons. These classes are further divided into subgroups with unique structural and functional characteristics, demonstrating the significant variability among these elements. Despite this variability, the fruit fly and other eukaryotic organisms use conserved mechanisms to regulate transposons. We will focus in this review on the transposition mechanisms and regulatory pathways of transposable elements as well as their functional roles in *Drosophila melanogaster*.

Transposons are also classified as autonomous or non-autonomous, depending on whether they transpose independently or require the machinery of autonomous transposons for mobilization.

As transposons comprise more than 40% of the human genome and are linked to numerous diseases, understanding their mechanisms of mobilization and regulation is important. *Drosophila melanogaster* is an ideal model organism for the study of eukaryotic transposons as its very well-known genome contains a diverse array of active transposons.

2. DNA TRANSPOSONS

DNA transposons are also called terminal inverted repeat (TIR) transposons. They consist of a transposase gene flanked by TIRs and move via a cut-and-paste mechanism. TIRs are repeating sequences found at both ends of these elements, and are inverted with respect to each other. The transposase is responsible for excising the transposon and inserting it into a new location. The regulation of DNA transposons in somatic cells is poorly understood, though some regulatory mechanisms have been identified for P elements in *D. melanogaster*. DNA transposons are divided into 2 sub-classes based on their transposition mechanisms.

Sub-class I elements use the canonical cut-and-paste mechanism of transposition and are divided into several superfamilies: Tc1/mariner, PIF/Harbinger, hAT, Mutator, Merlin, Transib, P, piggyBac, and CACTA. They can be autonomous or non-autonomous.

P elements

P elements are the best-studied DNA transposons in the *D. melanogaster* genome. Full-length autonomous P elements are 2.9 kb in length with 31 bp TIRs. Like other TIR transposons, P elements use a cut-and-paste mechanism of transposition. The complete, autonomous P element encodes an active transposase enzyme through splicing.

Transposition is initiated by interactions between direct repeats of DNA in the TIRs and the transposase (**Figure 1**). The dimerization of TIR-bound transposases induces the cleavage of the element. The excised element can be inserted into a new target site and through gap repair will generate target site duplications. Target site duplications are characteristic of TIR transposon insertions and may be used to identify transposition events and distinguish between different families of TIR transposons. For example, P element insertion and subsequent excision results in the production of 8 bp direct repeats.

P elements were discovered in *Drosophila* as they play a significant role in hybrid dysgenesis syndrome, a phenomenon observed in the progeny of hybrid crosses of certain wild-type *Drosophila* strains. However, P elements are also found in many other eukaryotes.

The regulation of P elements in *D. melanogaster* is better understood than that of other DNA transposons. P element transposition is regulated primarily by alternative splicing of the P element transposase mRNA. In germline cells, all 3 introns are spliced out, producing a functional transposase. Alternatively, in somatic cells, splicing of the third intron is skipped, generating mRNA that encodes a non-functional protein due to an early stop codon in the third intron. The resulting truncated transposase is not only inactive but represses the transposition of P elements in somatic cells.

In P strains, autonomous P elements are abundant while in M strains there are no P elements. P strains encode a P element repressor which accumulates in the cytoplasm during the development of cells. As the cytoplasmic conditions of the P cytotype are exclusively transmitted maternally, crosses between P type female with M or P type males do not result in hybrid dysgenesis because the repressors will bind to any P elements preventing their transposition (**Figure 2**). On the other hand, M type females, which lack the P cytotype, and P type males, which cannot pass on their P cytotype, result in hybrid dysgenesis syndrome due to an unregulated P element mobilization in the germlines of the hybrid progeny. The phenomena observed in these F1 hybrids include high rates of mutation, recombination, and sterility.

DNA transposon regulation generally appears to occur by generation of a non-functional transposase, which may include alternative splicing mechanisms. P elements are commonly used as mutagenic agents in genetic experiments.

There are also non-autonomous P elements which contain an internal deletion of varying length preventing transposase production, but such elements can still be mobilized if transposase is encoded elsewhere in the genome. We can find many other non-autonomous DNA transposons in eukaryotes like miniature inverted repeat transposable elements (MITEs)

Sub-class II DNA transposons include Helitron and Maverick elements that utilize unique transposition mechanisms.

Helitrons

Unlike other DNA transposons, Helitrons lack TIRs and encode a DNA helicase and replicator initiator (Rep) protein with nuclease and ligase functions. *Drosophila*

interspersed nuclear element-1 (DINE-1), the most abundant transposon in the *D. melanogaster* genome, is a non-autonomous Helentron, a subclass of Helitrons.

Helitrons use a rolling-circle replication mechanism of transposition (**Figure 3**). Rep protein nick the 5' end of one Helitron strand at a conserved TC sequence and the AT sequence on the target site. The Helitron donor strand is displaced by the encoded helicase. Rep cleaves the donor strand at a conserved hairpin signal in the 3' end, generating a circular, single-stranded DNA (ssDNA) intermediate. To complete transposition of the element, Rep cleaves the circular ssDNA intermediate to promote covalent bond formation between the 5' and 3'ends of the Helitron donor strand and the nicked target site. Host DNA replication is responsible for generating the second strand at both the donor and target sites, permitting amplification of these elements.

3. RNA TRANSPOSONS

Retrotransposons, or RNA transposons, are the most abundant class of transposons in the *D. melanogaster* genome. Retrotransposons use a copy and paste mechanism by first generating an RNA intermediate that is then reverse transcribed by an encoded reverse transcriptase (RT) into a new DNA copy that is inserted elsewhere in the genome. They are classified as either long-terminal repeat (LTR) retrotransposons, non-LTR retrotransposons (LINEs) or short interspersed nuclear elements (SINEs), depending on the presence or absence of LTRs required for element mobilization.

Retrotransposons are primarily characterized by the presence of *gag* and *pol* genes. These genes resemble those of retroviral genomes in both structure and function as these transposable elements arose from retroviruses that lost infectivity, for example by losing the gene *env* that encoded for the viral envelope. The retrotransposon pol gene encodes a polyprotein, typically consisting of a protease involved in processing the polyprotein precursor, an integrase required for the insertion of cDNA into the host genome, and a reverse transcriptase (RT) required for reverse transcription of the RNA intermediate to generate DNA copies of these transposons. RT is, therefore, common to all autonomous retrotransposons. On the other hand, gag proteins provide a structural coat for components involved in the reverse transcription event of retrotransposon mobilization.

In *D. melanogaster*, the regulation of retrotransposons in somatic cells is mediated by siRNAs, which are generated by cleavage of long dsRNA precursors derived from convergent sense and antisense transcription of retrotransposons in the genome. It has been seen that these siRNAs regulate retrotransposons in the nucleus via induction of heterochromatin formation. A similar pathway has been reported in humans to regulate LINE-1 retrotransposons via interfering RNAi.

LTR retrotransposons

In *D. melanogaster*, there are 3 recognized groups of LTR retrotransposons (Gypsy, Copia, and BEL/Pao).

LTRs play a significant functional role in the mobilization of these elements. Typically, LTR-retrotransposon mRNAs are produced by the host RNA pol II. The same mRNA molecule can be both translated into Gag and Pol proteins in the cytoplasm or captured

in a virus like particle (VLP) to use it as the reverse transcription template. The Gag and Pol genes are encoded in the same mRNA. Depending on the species, two different strategies can be used to express the two polyproteins, either a fusion into a single open reading frame (ORF) that is then cleaved or the introduction of a frameshift between the two ORFs. Occasional ribosomal frameshifting allows the production of both proteins ensuring much more Gag protein content that will form virus like particles in the cytoplasm.

Reverse-transcription usually initiates at a short sequence located immediately downstream of the 5'-LTR called the primer binding site (PBS) (**Figure 4**). Specific host tRNAs bind to the PBS and act as primers for reverse-transcription, which occurs in a complex process (Figure) ultimately producing a double- stranded cDNA molecule. The cDNA is finally integrated into a new location and adding a new copy in the host genome.

While many retrotransposons demonstrate no specificity for target insertion sites, elements of the gypsy family of LTR retrotransposons in *D. melanogaster* show some target site preference. Due to their structure, LTRs also permit recombination events in regions of the genome with high recombination rates resulting in spontaneous mutations. That is why due to selection against these mutagenic events, the large majority of LTR retrotransposons are detected in the inaccessible heterochromatin of *Drosophila* chromosomes where we find low recombination rates. However, LTR retrotransposon insertions are not limited to heterochromatic regions, and may even occur in protein-coding regions of the genome. These are selected against over time.

Non-LTR retrotransposons

Non-LTR retrotransposons, or LINE-like elements, have been classified into 6 groups: R2, L1, RTE, I, Jockey and RandI. Non-LTR retrotransposons are structurally similar to LTR retrotransposons, but do not contain LTRs at their 3' and 5' ends and often lack some of the protein domains encoded by LTR retrotransposons. The absence of LTRs suggest they use different mechanisms of transposition.

Studies in *D. melanogaster* demonstrated that non-LTR retrotransposons use target primed reverse transcription (TPRT) for integration of new retrotransposon copies in the genome.

R2 elements transposition

R2 non-LTR retrotransposons are co-transcribed with their flanking 28S rRNA sequences so these elements target 28S rRNA genes for insertion. TPRT is initiated by single-stranded nicking of the target DNA by the element's encoded endonuclease (**Figure 5**). The generated 3' hydroxyl group then primes reverse transcription of the RNA intermediate before cleavage of the second target DNA strand. The RT/EN encoded by R2 is responsible for the cleavage of the target DNA and reverse transcription of the element. The 3' UTR of these elements is required for TPRT and is inserted during reverse transcription of the element. The R2 RT/EN also generates the complementary R2 strand at the target site to fully transpose the element. It has been proposed another model for R2 insertion where two R2 elements may use their flanking 28S rRNA sequences to bind regions of homologous chromosomes, generating a Holliday junction

structure that is resolved by a dimer of the elements' endonucleases. These mechanisms may also be used by other non-LTR retrotransposons.

A few closely related non-LTR retrotransposons present in all *Drosophila* genomes (HeT A, TART, and TAHRE) play a significant role in telomere maintenance and use a unique mechanism to localize during transposition. These elements are targeted to telomeres by their encoded Gag proteins, permitting the generation of telomeric tandem repeats and performing functions similar to that of telomerase.

4. DISCUSSION

The mechanisms by which transposons mobilize in *Drosophila* and other eukaryotic genomes reveal several common features. For example, the enzymes encoded by these elements: transposases, integrases and endonucleases that permit the cleavage of target sites in the host genome and promote insertion of these elements into these genomic locations. These enzymes also require interactions with specific structural elements that flank the ORFs of the transposons such as TIRs in DNA transposons or LTRs, and 3' or 5' UTRs in RNA transposons.

Furthermore, many transposons have demonstrated the ability to amplify upon mobilization through RNA intermediates in the case of retrotransposons. Similarities between LTR and non-LTR retrotransposon mobilization are also evident, such as the formation of VLPs in the cytoplasm via polymerization of encoded Gag proteins, an event resembling a stage in the retroviral life cycle.

In addition to similar mechanisms of transposition, transposons are regulated by a common mechanism in *D. melanogaster*: small RNA synthesis. Regulatory pathways rely on convergent sense and antisense transcription of transposons to generate the precursors of the siRNAs.

The factors that regulate mobilization of transposons in host genomes may significantly influence genome size evolution, as transposon abundance correlates with genome size and this depends on the efficiency of transposition regulation and selection against deleterious transposition events over time.

Functionally, transposons may have a broad range of impacts on their hosts. Most deleterious integrations of transposons into host genomes are negatively selected against over time, while some insertions may provide adaptive functions to their hosts, such as the insertion of the solo-LTR FBti0019985 which acts as a cold stress response gene in *D. melanogaster*. Furthermore, the role of transposons in *Drosophila* telomere maintenance demonstrates the ability of these elements to develop significant functional roles to positively influence genome stability.

All of this has strong implications for the role of transposons in the evolutionary development of host genomes, as selective forces act on these transposition events, influencing the coevolution of the genome and its transposable elements.

5. BIBLIOGRAPHY

Chuong, E. B., Elde, N. C., & Feschotte, C. (2016). Regulatory activities of transposable elements: from conflicts to benefits. Nature Reviews Genetics, 18(2), 71-86. https://doi.org/10.1038/nrg.2016.139

Levin, H. L., & Moran, J. V. (2011). Dynamic interactions between transposable elements and their hosts. Nature Reviews Genetics, 12(9), 615-627. https://doi.org/10.1038/nrg3030

McCullers, T. J., & Steiniger, M. (2017). Transposable elements in Drosophila. Mobile Genetic Elements, 7(3), 1-18. https://doi.org/10.1080/2159256x.2017.1318201

Wikipedia contributors. (2020, abril 15). LTR retrotransposon. Recuperado de https://en.wikipedia.org/wiki/LTR_retrotransposon

FIGURES

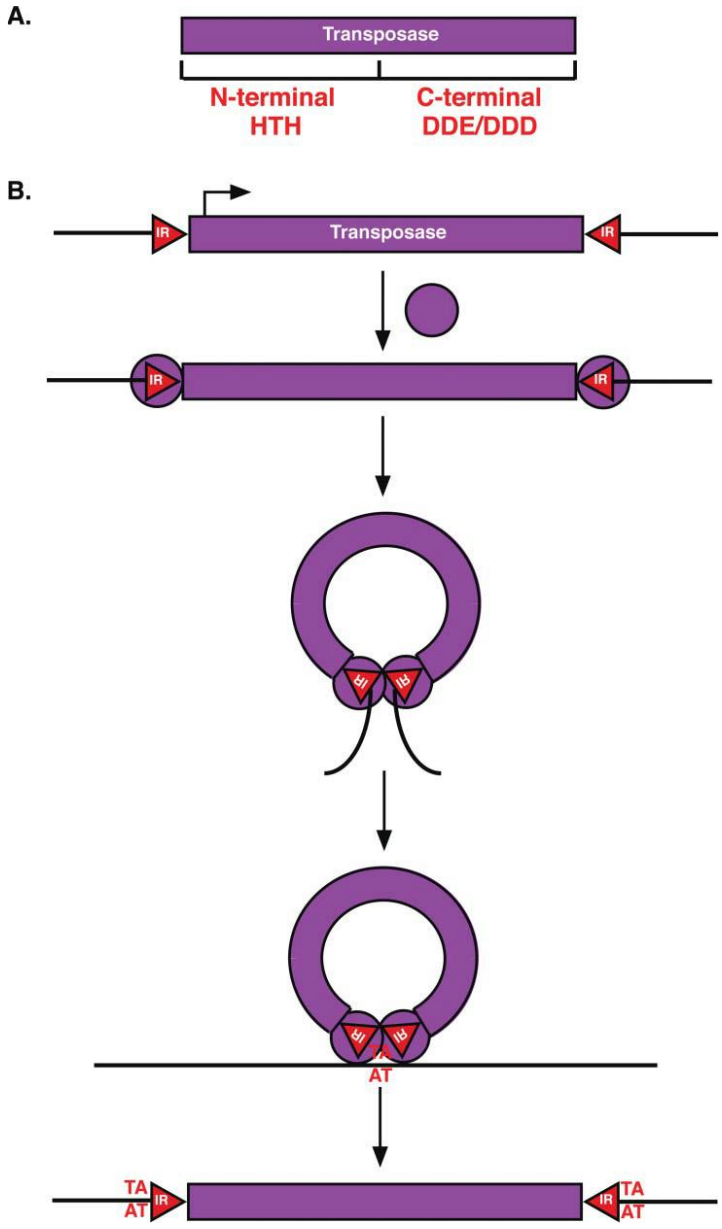

Figure 1: TIR transposase and transposition mechanism. (A) TIR Transposases have an N-terminal DNA binding domain with HTH motifs and a C-terminal DDE or DDD catalytic domain. (B) For transposition, TIR transposases (purple circles) first bind to inverted repeats (red triangles, IR) flanking the element. Bound transposases then dimerize followed by cleavage of the element from surrounding sequences (black lines) and integration into a new target site (AT) resulting in target site duplication.

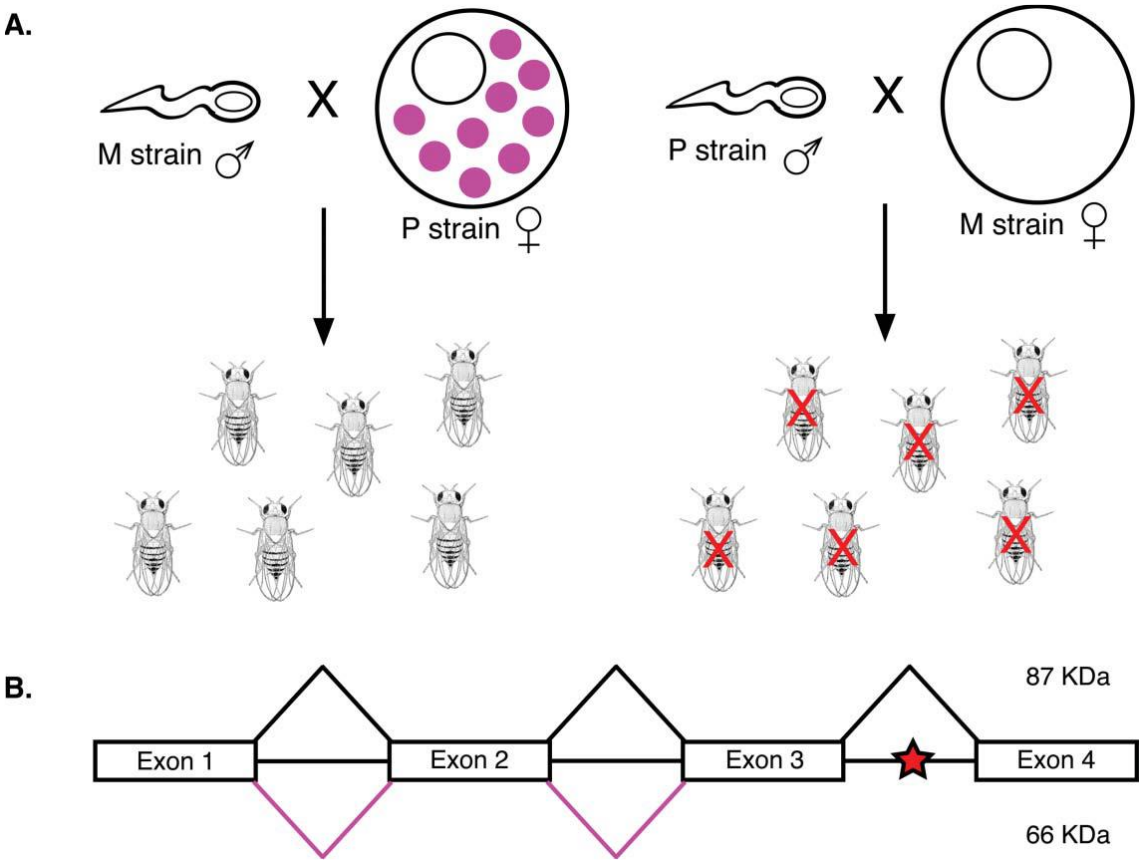

Figure 2: P element splicing and hybrid dysgenesis. (A) Hybrid dysgenesis results when M strain females are crossed with P strain males. Because the P element repressor (pink circles) is only transmitted by P cytotype females, progeny of the P strain male-M strain female cross have many mutations caused by germline P element transposition. These mutations often result in sterility (red X). (B) Exons 1-4 of P element transcripts are spliced to form a functional 87 kDa transposase (black lines). When intron 3 is not properly spliced, a stop codon (red star) generates a 66 kDa truncated repressor of P element transposition (pink lines).

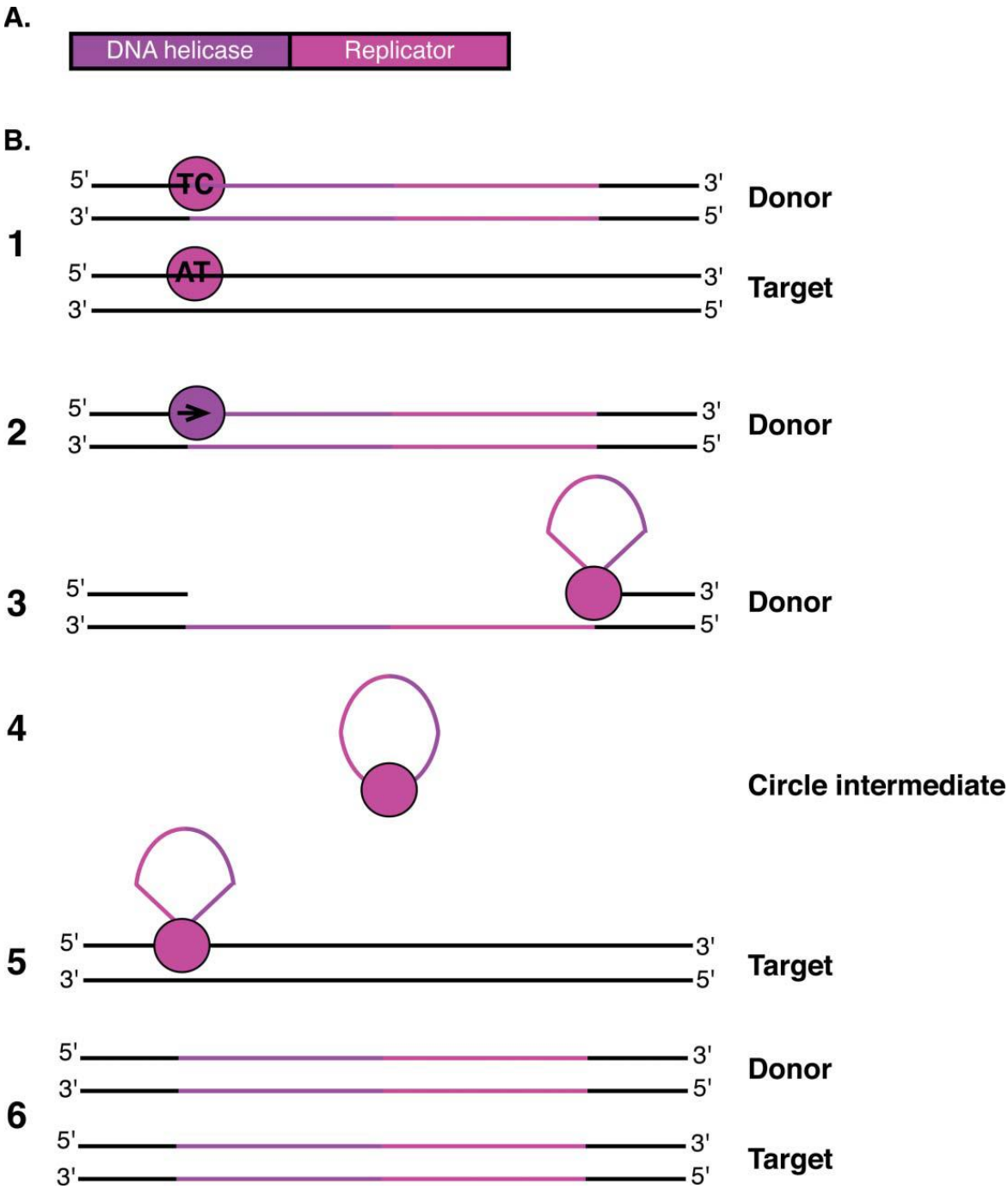

Figure 3: Helitron enzymes and transposition mechanism. (A) Helitron transposons encode a protein with both DNA helicase and Replicator functions. (B) The Helitron is represented with purple and pink lines.**1** The Replicator domain (pink circle) first binds to both donor (TC) and target (AT) creating nicks in both.**2** The DNA helicase domain (purple circle) then displaces the donor strand.**3** The Replicator domain cleaves the 30 end of the element, promoting formation of a circular single-stranded DNA intermediate.**4** Rep cleaves the circular single-stranded intermediate and promotes covalent bond formation between the 5' and 3' ends of the donor strand and target site.**5** Host DNA replication generates a second DNA strand at both the donor and target sites.**6** While the Replicator nicks the other end of the donor and facilitates attachment to the target site. The second strand of the element is generated at both the donor and target sites upon host DNA replication.

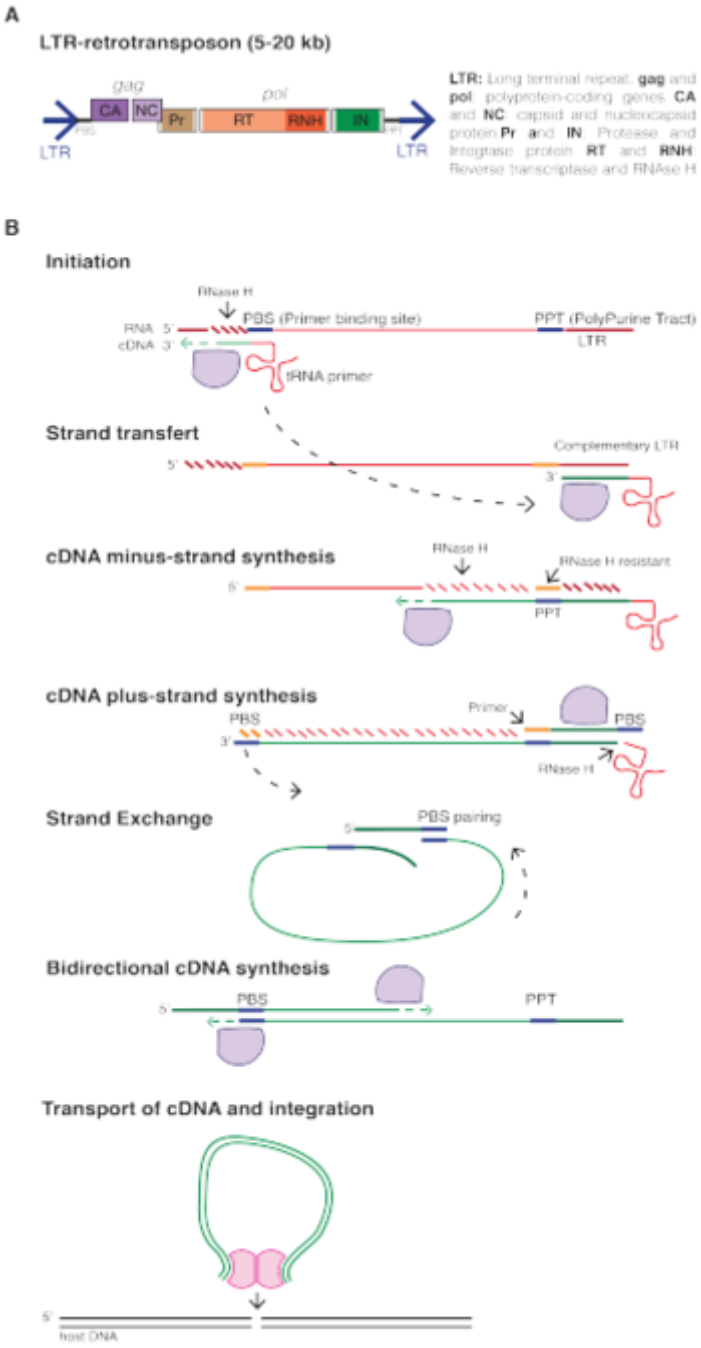

Figure 4. LTR-retrotransposon transposition. A. Genetic structure of LTR-retrotransposons (gypsy-type). B. Mechanism of retrotransposition, occurring inside viral-like particles in the cytoplasm. Reverse-transcription initiates at a host tRNA primer binding site (PBS) located immediately downstream of the 5'LTR. The newly synthesized minus-strand cDNA copy of the 5'LTR is then transferred to the 3'LTR and used as a primer for reverse-transcription of the entire minus-strand sequence. An RNase H-resistant polypurine tract then serves as a primer for plus-strand synthesis of the 3'LTR and complementary PBS. The newly-synthesized plus-strand PBS then associates with the already-synthesized minus-strand PBS, and double-stranded cDNA is finally produced. Double-stranded cDNA is then transferred to the nucleus by integrase proteins, and a new copy is integrated into the genome.

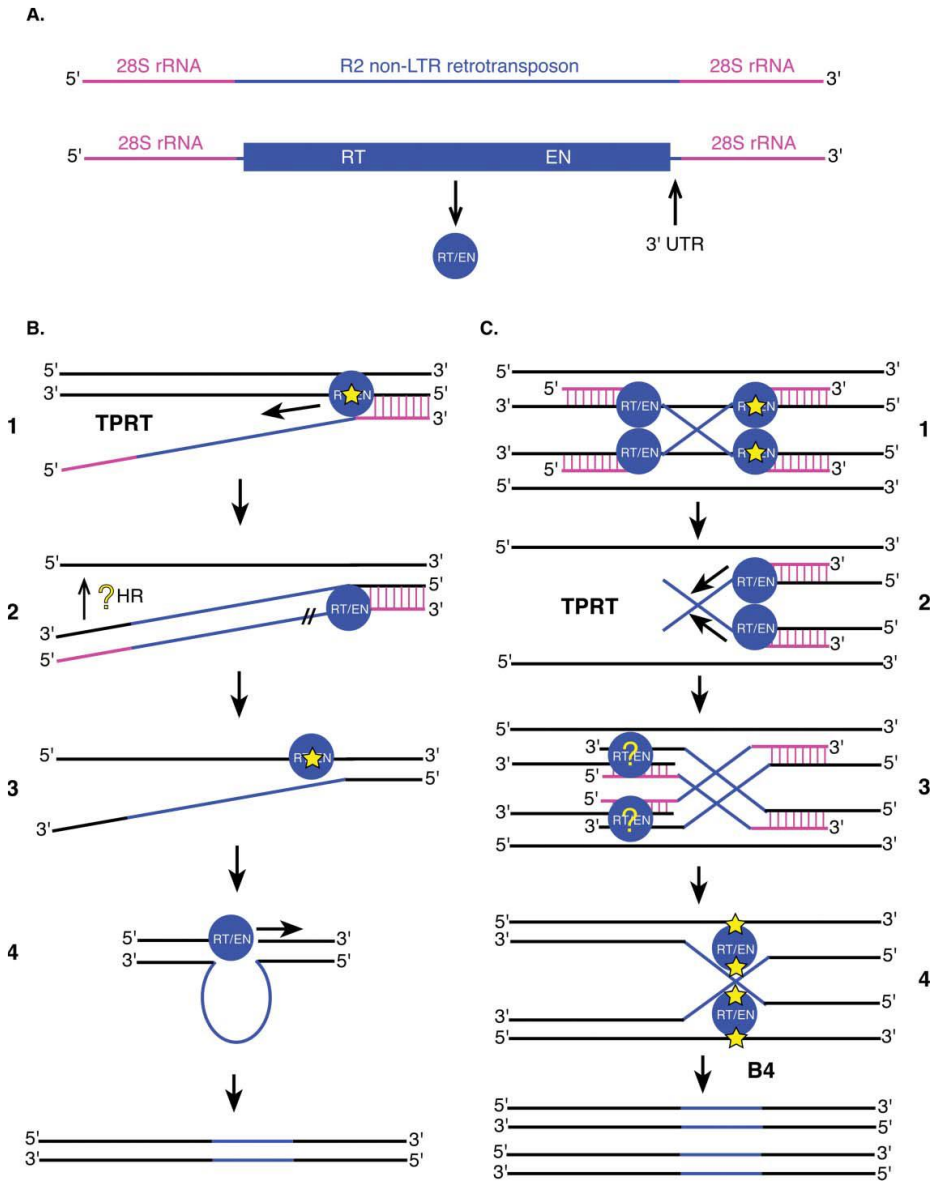

Figure 5: Non-LTR retrotransposons utilize target primed reverse transcription (TPRT) for integration. (A) R2 non-LTR retrotransposons (blue) are flanked by 28S rRNA genomic sequences (pink). The single R2 ORF encodes an enzyme with reverse transcriptase (RT) and endonuclease (EN) activities (blue circle). Other non-LTR retrotransposons may encode these enzymes as 2 separate proteins (RT and integrase with EN activity). The 30 UTR is important for integration of R2 retrotransposons into 28S rRNA genes. (B and C) Proposed models of non-LTR retrotransposon (B) and R2 (C) insertion. DNA is shown in black (including reverse transcribed flanking sequences), 28S rRNA sequences in pink and retrotransposon sequences (mRNA and DNA) in blue, following the color scheme in (A). (B1) Non-LTR retrotransposon transcripts first hybridize to 28S rRNA sequences (vertical pink lines) followed by initiation of TPRT by single-stranded nicking of the target DNA (yellow star) by the element's encoded endonuclease. (B2) Following reverse transcription of the element, element mRNA is degraded by R2 RT/EN (//). Integration of the 50 end of the element is not well understood (yellow ?). (B3) Cleavage of the second strand (yellow star) may occur at the same location as the first strand, or 2 base pairs upstream or downstream of this site. (B4) R2 RT/EN also generates the complementary R2 strand at the target site to fully transpose the element. (C1, C2, C3) The initial steps of this alternative mechanism are identical to those described in B1 and B2 except they take place on 2 homologous targets simultaneously resulting in a Holliday junction intermediate (C4). The Holliday junction intermediate is resolved by R2 RT/EN (C4) followed by second strand synthesis resulting in fully-integrated R2 non-LTR retrotransposons in 2 new locations.