Journal of Genetics and Genomics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Genetics and Genomics

Journal homepage: www.journals.elsevier.com/journal-of-genetics-and-genomics/

Review

The power of "controllers": Transposon-mediated duplicated genes evolve towards neofunctionalization

Huijing Ma [a, 1], Mengxia Wang [a, b, 1], Yong E. Zhang [a, b, c, d, *], Shengjun Tan [a, *]

[a] Key Laboratory of Zoological Systematics and Evolution & State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China
[c] CAS Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, Yunnan 650223, China
[d] Chinese Institute for Brain Research, Beijing 102206, China

ARTICLE INFO

Article history:
Received 22 December 2022
Received in revised form
4 April 2023
Accepted 5 April 2023
Available online xxx

Keywords:
Transposable elements
Gene duplication
New gene origination
End bypass
Template switching
Exon shuffling
Neofunctionalization
Mutational constraint

ABSTRACT

Since the discovery of the first transposon by Dr. Barbara McClintock, the prevalence and diversity of transposable elements (TEs) have been gradually recognized. As fundamental genetic components, TEs drive organismal evolution not only by contributing functional sequences (*e.g.*, regulatory elements or "controllers" as phrased by Dr. McClintock) but also by shuffling genomic sequences. In the latter respect, TE-mediated gene duplications have contributed to the origination of new genes and attracted extensive interest. In response to the development of this field, we herein attempt to provide an overview of TE-mediated duplication by focusing on common rules emerging across duplications generated by different TE types. Specifically, despite the huge divergence of transposition machinery across TEs, we identify three common features of various TE-mediated duplication mechanisms, including end bypass, template switching, and recurrent transposition. These three features lead to one common functional outcome, namely, TE-mediated duplicates tend to be subjected to exon shuffling and neofunctionalization. Therefore, the intrinsic properties of the mutational mechanism constrain the evolutionary trajectories of these duplicates. We finally discuss the future of this field including an in-depth characterization of both the duplication mechanisms and functions of TE-mediated duplicates.

Introduction

As a subject worthy of a Nobel Prize award, transposable elements (TEs) were initially discovered by Dr. Barbara McClintock in the 1940s (McClintock, 1948). With 80 years of effort, the following three characteristics of TEs are well-known: 1) although TEs are extremely diverse, they can be roughly grouped as Class I TEs (retrotransposons) with RNA as transposition intermediates, and Class II TEs (DNA transposons) with DNA as intermediates (Finnegan, 1989; Wicker et al., 2007; Wells and Feschotte, 2020); 2) with very few exceptions, TEs constitute an appreciable proportion of almost all eukaryote genomes (Wicker et al., 2007; Sun et al., 2012; Wells and Feschotte, 2020); and 3) TEs shape the activity and evolution of host genomes through contributions of regulatory or coding sequences and through induction of genomic changes (Dooner and Weil, 2007; Cosby et al., 2019; Fueyo et al., 2022; Wang et al., 2022a). The last characteristic, the functional consequences of TEs, draws the most attention. On the one hand, Dr. McClintock noticed that TEs contributed regulatory elements and thus controlled the activity of flanking genes as "controlling elements" (McClintock, 1956). In parallel, the coding sequences of TEs could also be co-opted or domesticated (Makalowski et al., 1994; Cosby et al., 2021). On the other hand, Dr. McClintock also recognized TEs as mutagens catalyzing structural changes in genomes including duplications, deletions, and inversions (McClintock, 1950).

Among these various changes, TE-mediated gene duplication was extensively studied. Specifically, gene duplication has long been believed to be the major mechanism underlying new gene origination and driving phenotypic evolution (Ohno, 1970; Zhang, 2003; Long et al., 2013). Although the evolutionary fate of most duplicates is pseudogenization or gene loss, a small proportion are retained or

* Corresponding authors.
E-mail addresses: zhangyong@ioz.ac.cn (Y.E. Zhang), tanshengjun@ioz.ac.cn (S. Tan).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jgg.2023.04.003

Please cite this article as: H. Ma, M. Wang, Y.E. Zhang *et al.*, The power of "controllers": Transposon-mediated duplicated genes evolve towards neofunctionalization, Journal of Genetics and Genomics, https://doi.org/10.1016/j.jgg.2023.04.003

preserved in long-term evolution (Lynch and Conery, 2000; Innan and Kondrashov, 2010). A dozen models have been proposed to explain the retention of these duplicates, among which positive dosage (selection for higher dosages), subfunctionalization (ancestral function subdivided across copies), and neofunctionalization (one copy conferring a novel role) models are more often studied (Zhang, 2003; Qian and Zhang, 2008; Innan and Kondrashov, 2010). Interestingly, due to the jumping and shuffling nature of TEs, TE-mediated duplicates tend to be neofunctionalized through new exon/intron structures or new regulatory contexts (Lynch and Katju, 2004; Katju, 2012; Cerbin and Jiang, 2018).

The first case of TE-mediated gene duplication was reported in bacteria as early as 1975 (Kleckner et al., 1975). Subsequently, tremendous efforts, including ours, have firmly established that gene or sequence duplication can be mediated by all major TE types (Fig. 1A). That is, genes can be captured by the two retrotransposon subclasses, long terminal repeat (LTR) and non-LTR retrotransposons (Moran et al., 1996; Emerson et al., 2004; Tan et al., 2016; Rosikiewicz et al., 2017; Zhang and Tautz, 2022), and by the two DNA transposon subclasses, terminal inverted repeat (TIR) TEs and rolling-circle TEs (also known as *Helitrons*) (Jiang et al., 2004; Thomas and Pritham, 2015; Tan et al., 2021). Given their ability in capturing and mobilizing host genes ("stealing" sequences from host genes), TEs are therefore dubbed as "thieves" (Lisch, 2005). Notably, due to the ubiquity and repetitive nature of TEs in genomes, they provide templates for non-allelic homologous recombination (NAHR) or ectopic recombination leading to gene excision or gene duplication (Babcock et al., 2003; Bailey et al., 2003; Kazazian, 2004; Yang et al., 2008). However, because NAHR is presumably enabled by all kinds of similar sequences, including TEs, it would not be discussed subsequently in this review.

Along with blooming research on TE-mediated gene duplication, numerous reviews have been published in the past decade. Some of them focused on one type of TE, such as duplicates captured by DNA transposons (Zhao et al., 2016), or retroduplicates (or retrocopies) generated by retrotransposons (Kubiak and Makalowska, 2017). Alternatively, some reviews broadly covered duplicates generated by various types of TEs, together with functions associated with these duplicates (Dooner and Weil, 2013; Cerbin and Jiang, 2018). Different from these previous efforts, we herein provide a mechanistic overview of duplicates generated by all four TE subclasses and then carefully dissect their common features. Furthermore, we discuss

A

	Subclass	Structure	Phylogenetic distribution of TEs	Example studies of TE-mediated duplicates
Class I	Non-LTR retrotransposon (*e.g.*, LINE-1)	ORF1 ORF2 (AAA)n	P, M, F, O	Human: 4611 (Rosikiewicz et al., 2017)
	LTR retrotransposon	gag pol	P, M, F, O	Fruit fly: 15 (Tan et al., 2016); Rice: 84 (Wang et al., 2006)
Class II	TIR TE	TPase	P, M, F, O	American alligator: 38 (Tan et al., 2021); Rice: 2967 (Zhao et al., 2018)
	Helitron	TC RepHel CTRR	P, M, F	Maize: 2694 (Dong et al., 2011)

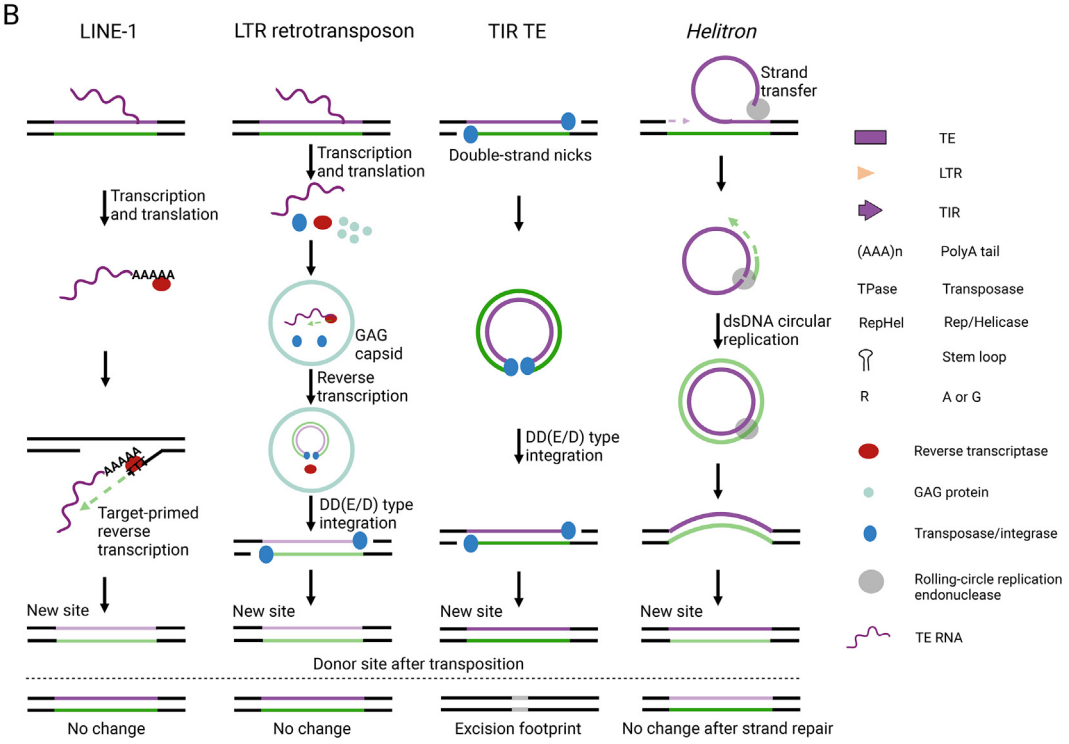


Fig. 1. The classification of TEs and their transposition mechanisms. **A**: TE classification system and representative studies of TE-mediated duplicates. Species groups P, M, F, and O are short for plants, metazoans, fungi, and others, respectively. In the last column, we listed one or two example studies with the number referring to the count of corresponding types of duplicates. **B**: The transposition mechanisms of four major TE subclasses. The purple and green solid lines indicate the double strand of a TE, while the dashed line indicates the newly synthesized strand. The figure was modified from Wells and Feschotte (2020).

how these commonalities predispose TE-mediated duplicates to the evolutionary trajectory of neofunctionalization. We end by providing our vision for the future of the field. Altogether, by diving into the dense literature on TE-mediated gene duplications, we highlight how the mechanistic properties of the mutational processes constrain evolution.

A concise overview of TE and TE-mediated gene duplication mechanisms

Among all four TE subclasses, retrocopies generated by non-LTR retrotransposons are the best studied, simply because one type of non-LTR retrotransposon, LINE-1s (also known as L1s), is the most active autonomous element in the human genome and its reverse transcriptase (RT) recognizes its own polyadenylated RNAs or host mRNAs (Fanning and Singer, 1987; Eickbush, 1992) via a target-primed reverse transcription (TPRT) process (Fig. 1B) (Luan et al., 1993; Moran et al., 1995). During the TPRT process, a nick is created at a loose site 5′-TTTT/A-3′ in a single strand of the target sequence with the endonuclease encoded by LINE-1 and the exposed 3′-OH is in turn used as a primer to initiate reverse transcription to synthesize cDNA of the LINE-1 mRNA under the catalyzation of reverse transcriptase (Feng et al., 1996; Cost et al., 2002). Then, the second strand of the target sequence is cleaved usually at 7−20 nucleotides downstream of the first nick (Kazazian and Moran, 1998) and the second strand is synthesized through RNA polymerase II or host repair mechanism to leave a target site duplication (TSD) of variable length flanking the insertio n. Later, we use LINE-1s to represent non-LTR retrotransposons. LINE-1-mediated retrocopies or retrogenes harbor all hallmarks of retroposition, including the absence of introns and the presence of a 3′ polyA tail or flanking TSDs.

LTR retrotransposons can also mediate the formation of retrocopies. These elements have long terminal repeats flanking the open reading frames (ORFs) and have a close evolutionary relationship with retroviruses (Wicker et al., 2007). LTR retrotransposons usually have at least two ORFs: one encodes the GAG protein, which is a structural protein forming virus-like particle (VLP); the other encodes the POL protein, which contains four domains: aspartic proteinase (AP), reverse transcriptase (RT), RH RNase H (RH) and DDE/D integrase (INT). The reverse transcription happens in the VLP, and then the cDNA bound by the integrase is integrated into host genome (Fig. 1B) (Wilhelm and Wilhelm, 2001; Delviks-Frankenberry et al., 2011; Wells and Feschotte, 2020). LTR retrotransposons are known to sometimes capture host mRNAs in their VLPs (Hajjar and Linial, 1993). Because LTR RT has low template affinity and low processivity, it may accidentally switch to host mRNA as a template (Goodrich and Duesberg, 1990) and then generate a chimeric cDNA. This mechanism is deeply conserved across plants (Wang et al., 2006, 2022b; Elrouby and Bureau, 2010; Hu et al., 2022), fungi (Derr et al., 1991; Schacherer et al., 2004), and animals (Jamain et al., 2001; Tan et al., 2016; Calatrava et al., 2022).

As for DNA transposons, both TIR TEs and *Helitrons* can duplicate host genes, which are called Pack-TIRs (“packing” duplicates into TIR TEs to form chimeric elements) and Pack-*Helitrons* (“packing” duplicates into *Helitrons*), respectively. TIR TEs use the canonical “cut-and-paste” mechanism for transposition (Fig. 1B). The transposase encoded by TIR TEs recognizes the TIRs and catalyzes the excision of both strands of the element from the donor site. Then the element is inserted into a distinct genomic position, where it is flanked by TSDs. This TE subclass consists of 23 superfamilies to date (Kojima et al., 2023). Among various TIR TE superfamilies, *Mutator*-like transposable elements (MULEs) in rice (Jiang et al., 2004) and CACTA elements (named by their conservative terminal inverted sequence “CACTA/G…C/TAGTG”) in sorghum (Paterson et al., 2009) are the two most successful gene-capturing superfamilies. However, the mechanism to generate Pack-MULEs and Pack-CACTAs in plants is largely unknown (Cerbin and Jiang, 2018). In animals, *P* elements (named by *P* strains carrying these TEs [Rubin et al., 1982]) in *Drosophila* are long known to capture sequences using the gap-filling model (Gloor et al., 1991), in which external sequences are filled in upon repair of double-strand breaks (DSBs) induced by transposition or secondary structures of TEs (Yant and Kay, 2003; Izsvak et al., 2004; Zhao et al., 2016; Tan et al., 2021). Different from TIR TEs, *Helitrons* encode 5′-TC and CTRR-3′ (R being G or A) termini instead of TIRs and in the 3′ end, it contains a 16−20 bp hairpin structure as the terminator with 10−12 bp away from the terminus. *Helitron*s transpose by “peel-and-paste” mechanism with peeling off the positive DNA strand to form an extrachromosomal circular DNA and replication by rolling circle (Fig. 1B) (Grabundzija et al., 2016, 2018; Kosek et al., 2021; Peng et al., 2022). Over 90% of *Helitrons* in maize have captured gene fragments (Dong et al., 2011) and similar events (Pack-*Helitrons*) are found in silkworms and bats, albeit in smaller numbers (Pritham and Feschotte, 2007; Han et al., 2013; Thomas et al., 2014). The gap-filling process may also underlie *Helitron*-mediated duplication (Feschotte and Wessler, 2001; Kapitonov and Jurka, 2007; Thomas and Pritham, 2015).

Notably, the end bypass (also named 3′ transduction) model is also proposed for both retrotransposons and DNA transposons in which the 3′ end is bypassed until an accidental downstream termination signal occurs (Tsubota and Huong, 1991; Moran et al., 1999; Feschotte and Wessler, 2001). In this case, the 3′ downstream sequence is duplicated by co-transposition.

Three common features of TE-mediated gene duplication mechanisms

Given the huge between-class or between-subclass divergence of TEs, fundamental differences are expected between the duplication mechanisms. Surprisingly, three important features, the end bypass process, template switching, and recurrent transposition are shared in gene duplications mediated by all four major TE subclasses.

End bypass

The end bypass process depends on two conditions, transposition activity and a weak or aberrant termination signal together with a fortuitous downstream signal. Therefore, more frequent transductions are expected for TEs with a simple termination signal, a simple recognition sequence, and a simple transposition process (Fig. 2A). The literature is consistent with this expectation.

LINE-1s are the most efficient TEs for generating 3′ transductions, followed by *Helitrons* and the other TEs. LINE-1s use a weak transcription termination signal, *i.e.*, the polyA signal (Moran et al., 1999), and show a *cis* transposition preference with proteins predominantly transposing their own transcripts (Wei et al., 2001). Due to these two facts, more than 20% of LINE-1s are associated with transductions (Goodier et al., 2000; Pickeral et al., 2000). *Helitrons* are commonly reported to mediate gene duplications through the end bypass model (Kapitonov and Jurka, 2007; Lal et al., 2009; Thomas and Pritham, 2015). The downstream terminator can be provided by a non-TE sequence or by another 5′ truncated *Helitron* (Kapitonov and Jurka, 2007; Tempel et al., 2007; Thomas et al., 2014). The latter is also called alternative transposition (Gray, 2000; Wang et al., 2015). In addition, one experiment revealed that the frequency of transductions increased after deletion of the hairpin in the 3’ end, which functions as the transposition terminator of *Helitrons* (Grabundzija et al., 2016).

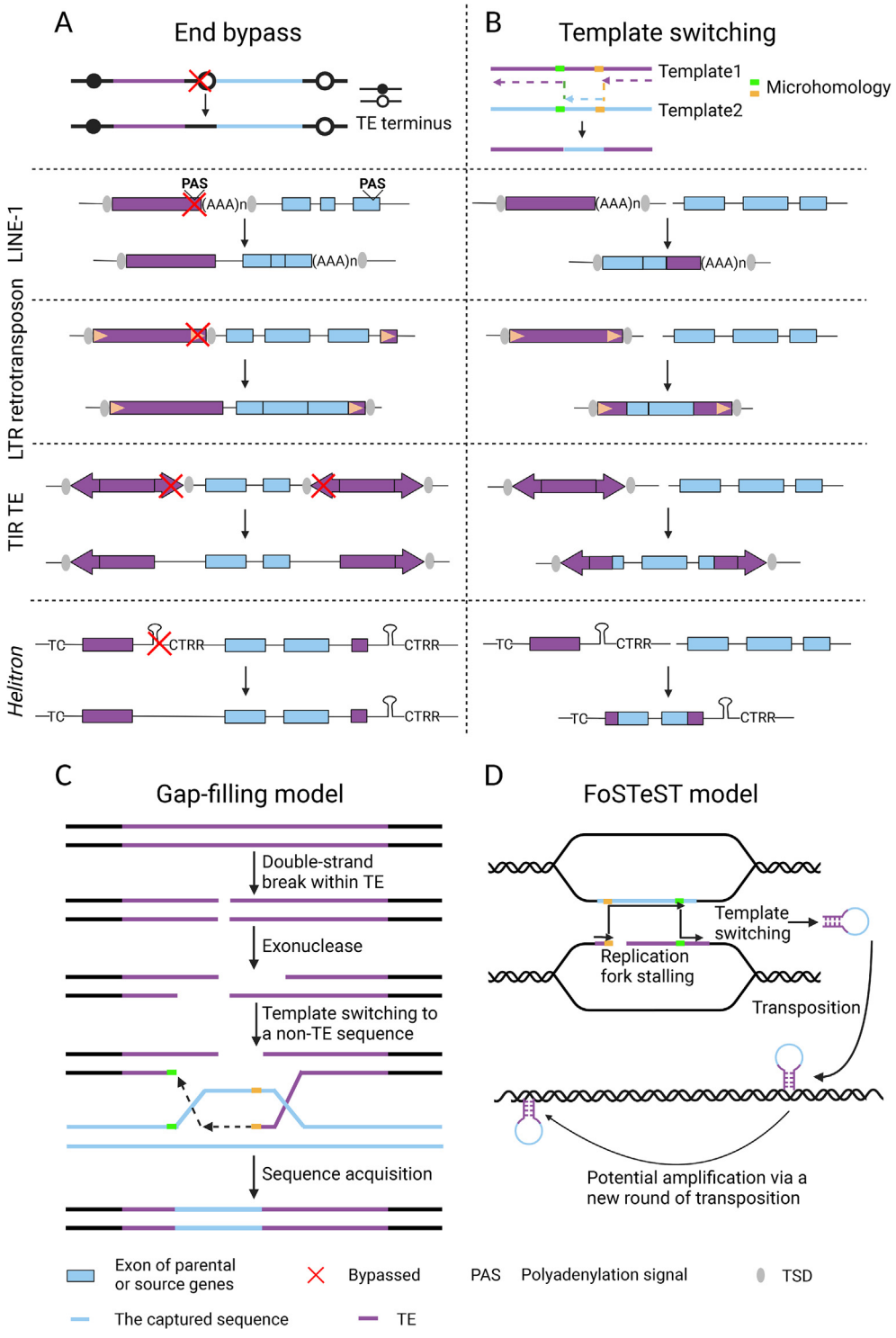


Fig. 2. The TE-mediated gene duplication mechanisms. For four major TE subclasses, TE-mediated gene duplications occur either via the end bypass or template switching mechanism. In the end bypass model (**A**), the termination signals in TEs are weak or aberrant, and the terminators of TEs are bypassed until an accidental terminator occurs. As a result, TEs with the 3′ flanking sequences are co-transposed to new sites. In the template switching model (**B**), microhomology (a short similar sequence shared by breakpoints of captured sequences and TEs) can induce polymerase or reverse transcriptase to switch from one template to another at the RNA or DNA level. After two rounds of template switches (or one round for LINE-1), the TE can capture a sequence from a parental gene (source gene) to form a chimeric element. In addition, two template switching models were further proposed for duplications mediated by TIR TEs: the gap-filling model (**C**) and the FoSTeST model (**D**). Panel **D** was adapted from Tan et al. (2021).

By contrast, the end bypass process is less likely used by LTR retrotransposons given their complicated life cycle (such as reverse transcription in VLPs) and the requirement for recognizing both LTRs. The *SUN* locus in tomato represents an outlier where readthrough of *Rider* led to duplication of *IQD12* (Xiao et al., 2008). Notably, several *Rider* readthrough transcripts were identified, suggesting that the termination signal was promiscuous or aberrant (Jiang et al., 2009). Similar to LTR retrotransposons, TIR TEs rarely induce transductions

because the terminal repeats are required to be bound by transposases. Only one anecdotal 3′ transduction event was documented in *Drosophila*, where an 8-bp sequence downstream of a *P* element was sufficient as the new TIR (Tsubota and Huong, 1991). A relatively more frequent mechanism is the alternative transposition model as in *Helitrons*. It is reported that two active CACTA elements can jump into genes and form closely spaced TE copies. The frequent deletion of TEs results in the loss of internal TIRs, and the gene together with the remaining external TIRs are mobilized as Pack-CACTAs (Catoni et al., 2019).

Template switching

Similar to the end bypass process, the template switching process happens at both the RNA and DNA level (Fig. 2B). For LINE-1s, reverse transcription starts from the 3′ end of their RNA templates and often aborts, thus causing 5′ truncation (Kazazian and Moran, 1998). It is reported that template switching could capture host sequences and attach them to the 5′ end of LINE-1s via microhomology (or microsimilarity, a short similar sequence shared by breakpoints of host sequences and LINE-1s) (Buzdin et al., 2002, 2003). Eighty-one chimeras consisting of different types of recombination (*e.g.*, U6 small nuclear RNA attached to LINE-1) have been detected in the human genome (Buzdin et al., 2003). Interestingly, 93% of the 5′ ends of these chimeras are from small nuclear RNAs involved in spliceosomes, indicating a close spatial distance between the reverse transcription and splicing machinery. Certainly, LINE-1s can also mediate retroposition by directly binding to the polyA tail of host mRNA rather than by template switching (Kaessmann et al., 2009). By contrast, for retroposition mediated by LTR retrotransposons, template switching is mandatory (Goodrich and Duesberg, 1990; Delviks-Frankenberry et al., 2011). Based on previous studies and ours, we conclude four features of template switching mediated by LTR retrotransposons (Tan et al., 2016): 1) the retrocopies tend to be partial compared to the parental transcripts; 2) both 5′ and 3′ LTRs usually coexist in the retrocopies, which is essential for recognition by the integrase; 3) microhomology at the fusion point between the mRNA and LTR retrotransposon is commonly identified, suggesting its importance in template switching; and 4) two or more switching events can occur, demonstrating that LTR-mediated retroposition has a strong ability to shuffle sequences.

For DNA transposons, the gap-filling model is essentially a template switching process (Fig. 2C). This model can be applied to both TIR- and *Helitron*-mediated duplications, which also has other names, such as synthesis dependent strand annealing (SDSA) or filler DNA model (Kapitonov and Jurka, 2007; Thomas and Pritham, 2015). Note that gap-filling only requires DSBs induced by excision via transposition or other mechanisms. Our recent work actually supports the transposition-independent gap-filling model by identifying duplicates flanked by TIR TEs (Pack-TIRs) in the primate genomes without active DNA transposons (Tan et al., 2021). This gap-filling model for DNA transposons and the LTR-mediated template switching model (Tan et al., 2016) are two similar processes occurring respectively at the DNA and RNA level, which share all aforementioned four features: partial copies, co-existence of 5′ and 3′ terminal repeats, microhomology, and sequence shuffling. We found that the gap-filling model is compatible with most single-copy Pack-TIRs, while we additionally proposed a replication Fork Stalling, Template Switching and Transposition (FoSTeST) model (Fig. 2D) to explain the generation of multicopy Pack-TIRs (Tan et al., 2021). The major difference between FoSTeST and the gap-filling model is that Pack-TIRs in the former model can be transposed as nonautonomous TEs and then amplified by recurrent transposition. In this aspect, FoSTeST is transposition-dependent.

Compared to end bypass, template switching shows two essential differences due to its distinct duplication mechanisms. First, template switching can happen between TEs and unlinked sequences. Second, it applies to scenarios with or without transposition activity (Tan et al., 2021). The latter scenario enables template switching as a more general mechanism of TE-mediated gene duplications at the DNA level.

Recurrent transposition

As exemplified by multicopy Pack-TIRs potentially generated by the FoSTeST process (Fig. 2D), duplicates generated by TEs can act as pseudo-TEs and amplify to multiple copies via transposition. The distinct copies are often dispersed on different chromosomes, which stands in contrast with the most frequent form of gene duplication induced by errors during DNA replication, recombination or repair, which is tandem duplications (Emerson et al., 2008; Zhang et al., 2022). Notably, TSDs help to distinguish transposition-mediated amplification from tandem duplications (Fig. 3).

In addition to Pack-TIRs, retrocopies mediated by LINE-1s or LTR retrotransposons can also experience two or more rounds of transposition (Zhang et al., 2003; Pan and Zhang, 2009; Tan et al., 2016; Batcher et al., 2020), as do Pack-*Helitrons* (Brunner et al., 2005; Morgante et al., 2005; Yang and Bennetzen, 2009; Dong et al., 2011). Clearly, as with *bona fide* TEs, the copy number of pseudo-TEs depends on the activity of the corresponding autonomous TEs. Two extraordinary examples come from the highly active *Helitrons* in bats: *HelibatN217.1*, in which a *Helitron* containing a fragment of *TMBIM4* (a gene involved in apoptosis inhibition) reached 392 copies (Thomas et al., 2014) and *HelibatN3* carrying part of nucleotide binding protein-like (*NUBPL*) was amplified to more than 1000 copies (Pritham and Feschotte, 2007).

Various evolutionary fates of TE-mediated duplicates

Like other types of duplicates, *e.g.*, tandem duplicates, the evolutionary fate of most TE-duplicates is also pseudogenization or gene loss. For example, the majority of LINE-1-mediated retrocopies were "dead-on-arrival" (Petrov et al., 1996; Zhang et al., 2003, 2004b)

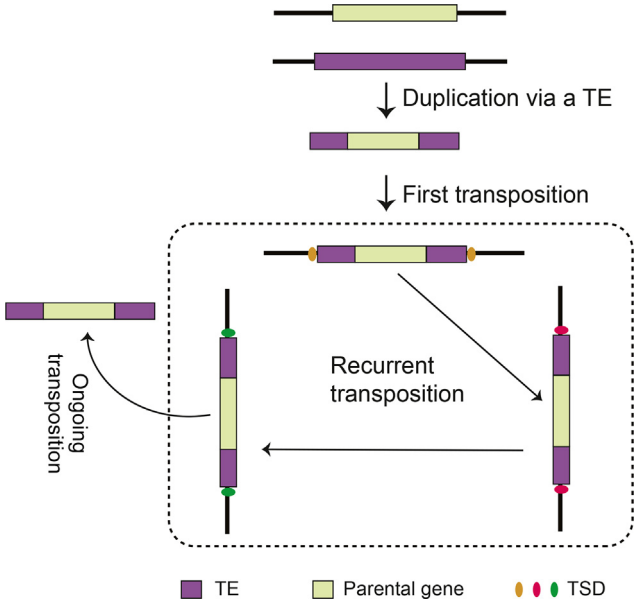


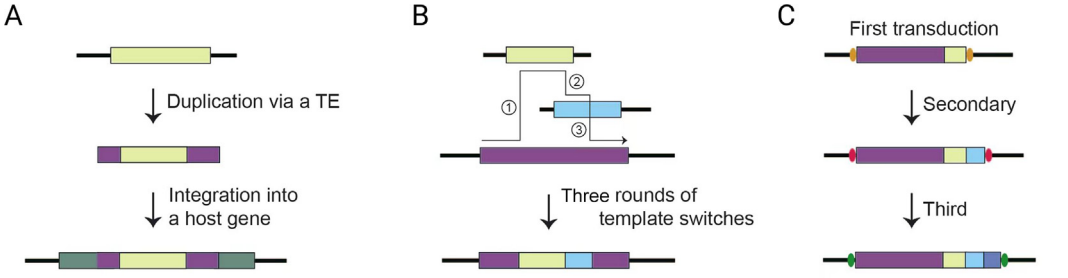
Fig. 3. Recurrent transposition. The duplicate generated by a TE may undergo recurrent transposition as a pseudo-TE.

due to the loss of preexisting promoters. Sometimes, TE-mediated duplicates can get transcribed using the nearby promoters provided by flanking TEs or the host genes. Therefore, TE-mediated duplicates can encode new proteins and evolve along varying trajectories including neofunctionalization, positive dosage and others. Although pseudogenization is the most common fate, we focus on the former evolutionary trajectories subsequently, which can explain the retention of the TE-mediated duplicates.

Exon shuffling contributes to neofunctionalization of TE-mediated duplicates

The aforementioned three common mechanistic features of the mutational process lead to one common functional outcome, exon shuffling (Fig. 4). Exon shuffling was originally proposed by Dr. Gilbert in 1978 in which exons represented the functional unit and were recombined as new genes via buffering of introns (Gilbert, 1978; Long et al., 2003; Zhang et al., 2022). Here, we used the term in a broader sense, in which exons, introns, or intergenic sequences are shuffled to generate novel gene structures (Batcher et al., 2022; Zhang et al., 2022). These chimeric sequences are expected to have novel functions compared to their source sequences (Betrán and Long, 2006; Xia et al., 2016; Zhou and Zhang, 2019).

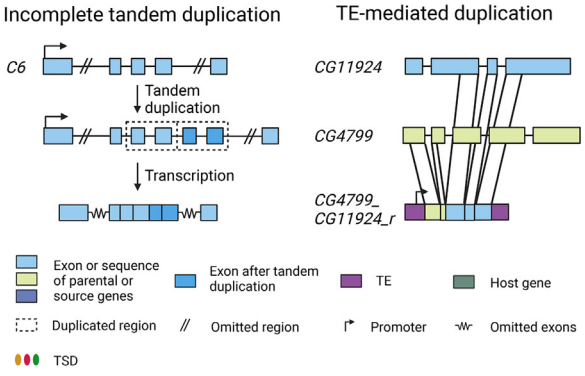
Two characteristics of TE-mediated duplicates may make them prone to exon shuffling. First, the mobile nature of transposition means that TE-mediated duplicates are shuffled into a new chromosomal context and expected to evolve under this context (Fig. 4A). One study showed that 17 out of the top 100 expressed retrocopies in humans were fused with the corresponding host genes (Vinckenbosch et al., 2006). The new gene, *jingwei* (*jgw*) in *Drosophila* is a classic example of neofunctionalization, which is contributed by exon shuffling. That is, *jgw* was shown to emerge via retroduplication of alcohol dehydrogenase (*Adh*) and then was fused with the host gene, *Yande* (Long and Langley, 1993). Subsequently, *jgw* was subjected to adaptive selection to gain a new function in substrate specificity (Zhang et al., 2004a, 2010). Similar neofunctionalization cases induced by retroposition and exon shuffling include *sphinx* (Wang et al., 2002) and *TRIM5-CypA* (Sayah et al., 2004) (Fig. 4D).



D

Species	Case studies	References
Vertebrates	*chiron*	Fang et al., 2021
	Cstf2t	Dass et al., 2007
	ENSMUSG00000021752_retro	Tan et al., 2016
	HAPSTR2	Amici et al., 2023
	PIP5K1A-PSMD4	Akiva et al., 2006
	PMSE2b	Zaiss and Kloetzel, 1999
	RBMXL1	Marques et al., 2005
	Super	Tan et al., 2016
	TRIM5-CypA	Sayah et al., 2004
	Utp14c	Rohozinski et al., 2006
Drosophila	*Adh-Twain*	Jones and Begun, 2005
	CG4799_CG11924_r	Tan et al., 2016
	CG7804	Lee et al., 2019
	Dntf-2r	Betrán and Long, 2003
	FB4-Ssk	Tan et al., 2021
	jingwei	Long and Langley, 1993
	Poseidon	Xia et al., 2022
	sphinx	Wang et al., 2002
Plants	*Bs1*	Elrouby and Bureau, 2010
	CYP98A8	Matsuno et al., 2009
	CYP98A9	

E



G

F

Species	Evidence	Tandem duplication	LTR or TIR TE-mediated duplication
D. melanogaster	Exon shuffling	111/270 (41.1%)[1]	15/15 (100%)[2]
	Chimeric transcripts	74/111 (66.7%)[1]	9/10 (90.0%)[2]
Human	Exon shuffling	196/964 (20.3%)[1]	8/8 (100%)[3]
	Chimeric transcripts	52/196 (26.5%)[1]	5/8 (62.5%)[3]

Fig. 4. Exon shuffling contributes to neofunctionalization of TE-mediated duplicates. **A**–**C**: Exon shuffling during TE-mediated gene duplications. In addition to scenarios with duplicates attached to TEs or captured by them (Fig. 2), there are three main forms of exon shuffling: integration into a preexisting host gene (**A**), several rounds of template switches in one transposition event (**B**), and multiple gene fragments captured step-by-step through recurrent transposition (**C**). **D**: TE-mediated duplicated genes with potential novel function. These cases are sorted by species and by gene symbols. A few representative genes are elaborated with more details in the text. Note that the novel functionality of certain genes has been experimentally established, while for others, it has been only inferred based on factors such as chimeric gene structure. **E**: Exon shuffling after incomplete tandem duplication in humans. In this example, intragenic partial duplication of *C6* leads to an extra copy of two exons. **F**: Statistics of exon shuffling cases in flies and humans associated with different duplication mechanisms. "Exon shuffling" indicates duplicates potentially involving shuffled sequences on the basis of sequence arrangement, while "Chimeric transcripts" indicates a subset of the former duplicates with transcriptome or RT-PCR evidence of chimerism. The number in parentheses indicates the proportion of exon shuffling genes. **G**: The formation of chimeric gene *CG4799_CG11924_r* in flies. Partial transcripts separately derived from *CG4799* and *CG11924* are co-captured by one LTR retrotransposon. (**E**) and [1] in (**F**) are from Zhang et al. (2022); (**G**) and [2] in (**F**) are from Tan et al. (2016), which analyzed duplications mediated by LTR retrotransposons; [3] in (**F**) is from Tan et al. (2021), which analyzed Pack-TIRs.

Strikingly, the fusion gene *TRIM5-CypA* emerged multiple times in different monkeys where *CypA* (binding the capsids of retrovirus) has experienced recurrent transposition into the same host gene *TRIM5* (involved in antiviral defense) (Kaessmann, 2009). *TRIM5-CypA* could more efficiently restrict retroviruses compared to the ancestral *TRIM5*. Second, TE-mediated duplicates tend to be chimeric upon birth when duplicated sequences are attached to TEs, as in LINE-1s, LTR retrotransposons and DNA transposons (Fig. 2). Thus, the duplicate may evolve a new gene structure with adjacent TEs or evolve under the regulatory sequences of TEs (Tan et al., 2016, 2021). Taking LTR-mediated retrocopies as an example, the majority (90%) of them are transcribed under the regulation of flanking transposons (Tan et al., 2016). From this perspective, either end bypass or template switching essentially provides a way to shuffle TEs and captured sequences, which can be further powered by recurrent transposition.

Template switching can enhance exon or sequence shuffling via multiple switches (Fig. 4B). Both LTR retrotransposons and DNA transposons can duplicate sequences from multiple genes. Previous work showed that retrocopies contain fragments from two to four genes in animals and plants. Cases include: *PIP5K1A-PSMD4* in hominoids, a retrogene derived from two neighboring genes (Akiva et al., 2006; Zhang et al., 2009); *Bs1* in maize, a retrogene fused from three different host genes (Elrouby and Bureau, 2010); a "*Super*" retrocopy derived from four genes in mouse; and a chimeric retrocopy *CG4799_CG11924_r* in the population of *D. melanogaster* (Fig. 4D) (Tan et al., 2016). For the latter two cases, their transcription is demonstrated. For *CG4799_CG11924_r*, we further detected reduced levels of nucleotide diversity and increased levels of linkage disequilibrium, suggesting that this gene is under positive selection and probably has evolved a new function (Tan et al., 2016). Certainly, in-depth experimental study is needed to reveal the functional roles of either "*Super*" or *CG4799_CG11924_r*. Notably, multiple switches seem to be more frequent for duplications mediated by DNA transposons in plants. About 20% of Pack-MULEs contain gene fragments from more than one locus in rice (Jiang et al., 2004), which are much more frequently expressed (32%) than those derived from one gene (21%) (Hanada et al., 2009), and about 70% of *Helitrons* capture more than one gene fragment in maize (Dong et al., 2011). Stepwise capture of gene fragments was detected in LINE-1s (Pickeral et al., 2000), MULEs (Jiang et al., 2004), and *Helitrons* (Dong et al., 2011; Thomas and Pritham, 2015), suggesting that recurrent transposition contributed to the formation of multigene chimeras (Fig. 4C). In contrast, whether these complex chimeras captured by LTR retrotransposons are generated by one mutational event or multiple sequential events remain unknown.

A natural question is whether TE-mediated gene duplication is more often subjected to exon shuffling compared to the most common type of gene duplication, *i.e.*, tandem duplication. So far as we know, there is no systematic study comparing these two types of gene duplications. However, we could make a preliminary inference on the basis of previous literatures. Specifically, we recently reported that incomplete tandem duplication is often subjected to exon shuffling by evolving new exon/intron structures (Fig. 4E) (Zhang et al., 2022). That is, 111 out of 270 (41.1%) polymorphic tandem duplications in flies are partial (spanning some exons rather than the whole genes), while the number becomes 196 out of 964 (20.3%) for humans (Fig. 4F). Transcriptional analyses demonstrate that 66.7% of partial duplicates in flies and 26.5% of partial duplicates in human evolve novel gene structures. Since the transcriptome data used in this study may be not complete, the upper-bound estimation of exon shuffling for tandem duplications in these two species would be 41.1% and 20.3%, respectively. By contrast, 9 out of 10 (90%) polymorphic LTR-mediated retrocopies evolve chimeric structures in flies (Fig. 4F and 4G) (Tan et al., 2016), so do 5 out of 8 (62.5%) recently originated Pack-TIR in humans (Tan et al., 2021). Therefore, the frequency of exon shuffling or chimerism seems to be higher for TE-mediated duplicates than for tandem duplications.

Notably, although neofunctionalization is anticipated to be the common fate of chimeric TE-mediated duplicates, the in-depth demonstration of neofunctionalization is not trivial and thus only a handful of case studies were performed with more or less functional evidence (Fig. 4D).

TE-mediated duplicates may also be retained due to a preference for a higher dosage or other reasons

For retained duplicates generated by all four TE subclasses, neofunctionalization is expected to be the most common fate (Fig. 5). However, although retrocopies mediated by LINE-1s are also subjected to possible neofunctionalization due to exon shuffling induced by 3′ transduction, attachment of 5′ small nuclear RNA, fusion with host genes (*e.g.*, *jgw* or *TRIM5-CypA*) or regulation by the host context, most retrocopies are derived from a single gene. Therefore, these retrocopies may also evolve under a dosage or subfunctionalization model. As aforementioned, although LINE-1-mediated retrocopies were generally "dead-on-arrival", a significant number of retrocopies have evolved into *bona fide* genes (*e.g.*, 117 retrogenes in humans), in which their transcription is often driven by nearby regulatory elements or permissive transcriptional environments (*e.g.*, testis, [Vinckenbosch et al., 2006; Kaessmann et al., 2009]). Such fortuitous expression is expected to be low or tissue-specific (*e.g.*, testis-specific). Thus, the chance for functional subdivision (subfunctionalization) on the basis of differential expression sites should not be high. By contrast, quite a few studies demonstrated that retrocopies could evolve under a dosage model (Fig. 5) (Parker et al., 2009; Sulak et al., 2016; Yang et al., 2020). For example, several retrocopies of *TP53* (*TP53RTGs*) are transcribed in elephants and contribute to the enhanced DNA damage response by increasing the *TP53* copy number (Sulak et al., 2016). In very rare scenarios, retrogenes may be blessed by a regulatory context more appropriate compared to that of parental genes, in which they evolve to be an orphan duplicate with the loss of corresponding parental genes (Ciomborowska et al., 2013; Carelli et al., 2016). Interestingly, we recently found that this unusual switch between the derived and progenitor copy also occurs in one dispersed segmental duplication presumably generated by recombination (Ma et al., 2022).

In this regard, although duplicates captured by LTR retrotransposons and DNA transposons may form chimeric sequences with flanking TEs and possibly evolve under the neofunctionalization model (Barbaglia et al., 2012; Han et al., 2013; Tan et al., 2016, 2021), sometimes they may also maintain their original gene structure and evolve under the dosage or subfunctionalization model (Fig. 5). We are not aware of cases of subfunctionalization, but the aforementioned *IQD12* duplication fits the dosage model in which the increased expression level led to elongated fruit shape (Xiao et al., 2008). Certainly, even if the duplicates maintain their preexisting gene structure, they could still evolve a novel function. As shown by one multicopy Pack-TIR family fitting the FoSTeST model, *Ssk-FB4s* (*Ssk* captured by *FB4*, with Ssk involved in smooth septate junctions of *Drosophila*), the amino acid substitution pattern and protein interaction data indicate a neofunctionalization consequence (Tan et al., 2021).

Significance and perspective

By surveying decades of research, we painted a picture where three common mechanistic features of TE-mediated duplications, including end bypass, template switching, and recurrent transposition, led to one common functional outcome, in which the

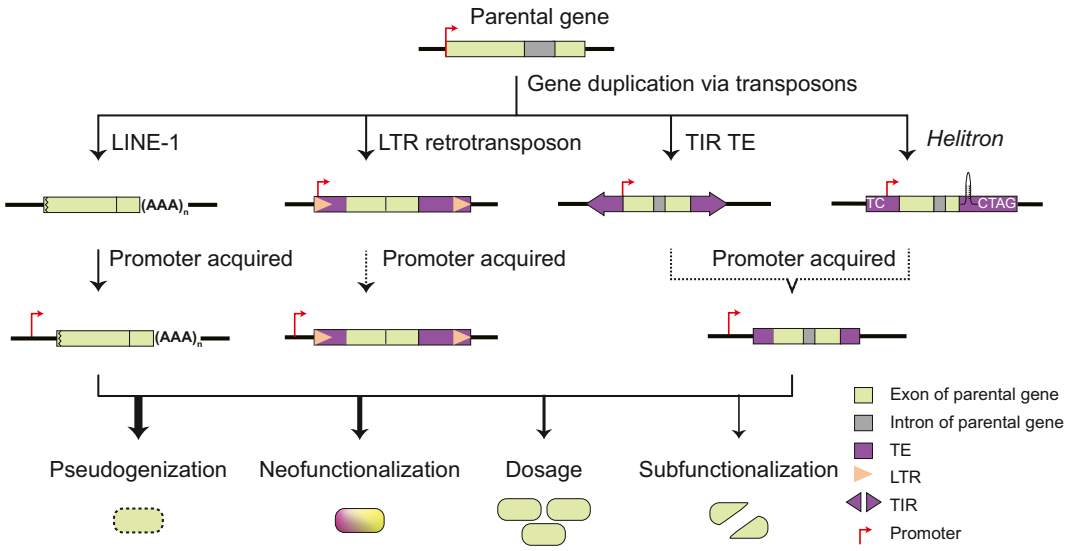


Fig. 5. Evolutionary trajectories of transposon-mediated duplicates. Duplicates shuffled into a new locus are expressed under a different regulatory context and may evolve along various trajectories. The red arrows mark the transcription start sites (TSSs) and the dashed lines indicate that these steps could be dispensable. In other words, the transcription of duplicates captured by LTR retrotransposons or DNA TEs could be initiated by the flanking TEs. The zigzag rectangle of LINE-1 indicates the retrocopy maybe 5′ truncated. For the four arrows near the bottom, the width indicates the relative prevalence of the corresponding evolutionary trajectories and the different shapes at the bottom represent the four fates of duplicates.

derived duplicates generally undergo exon shuffling and thus are expected to evolve in a trajectory which is different from that of the parental genes.

The significance of this unified picture is two-fold. On the one hand, TEs are generally perceived as extremely diverse given their divergent transposition machinery, which is shaped by an endless arms race between these molecular "parasites" and host genomes (Dooner and Weil, 2007; Cosby et al., 2019; Almeida et al., 2022). It is actually counter-intuitive that duplications mediated by the four TE subclasses share so many features in terms of mutational processes and functional consequences. A unified picture highlighting commonality provides a simple framework to grasp the individually complicated TE-mediated duplication processes. On the other hand, although it is undoubtedly true that nothing in biology makes sense except in the light of evolution (Dobzhansky, 1973); the parallel view, nothing in evolution makes sense except in the light of biology, is becoming appreciated in the recent decade (Nei, 2007; Rosenberg and Queitsch, 2014; Witt et al., 2019; Zhang et al., 2022). Supporting evidence for the latter view is especially pronounced in studies of mutational properties. For example, the recurrent enhancer deletion of pituitary homeobox transcription factor 1 (*Pitx1*) led to the parallel evolution of pelvic reduction in three-spine stickleback fish populations, through contributions by both deletion bias (a magnitude higher deletion rate relative to the genomic background) and adaptive selection (Chan et al., 2010; Xie et al., 2019). TE-mediated duplication studies fit the same line of thinking, in which the derived copies generated by TEs are born with gene structures different from their progenitors, and their future evolution toward neofunctionalization is guided by mutational properties.

Despite significant progress in the past decades, two fundamental questions of TE-mediated gene duplications, mutational mechanisms and the function of duplicate genes, still warrant future exploration. Due to a strong interest in LINE-1s in mammals, the corresponding retroposition mechanism is extensively studied (Moran et al., 1999; Wei et al., 2001; Cost et al., 2002; Hwang et al., 2021). The mechanisms for other TE subclasses are at best partially characterized. Using LTR-mediated retroposition as an example, we previously observed that the RT of *Gypsy BLASTOPIA* always switches back from the mRNA to the retrotransposon at one particular site (TAAAAAACAAGC-GGTTG, [Tan et al., 2016]). This phenomenon raises two questions, namely, how the switch between the LTR retrotransposon and mRNA is regulated, and whether such regulation is shared across TE families. With such questions addressed, we will not only know which genes are more prone to be captured by transposons and thus have higher duplicability, but will also be able to select better RTs with preferred switching behavior (*e.g.*, low switching frequency) for bioengineering (Martin-Alonso et al., 2021). As for the function of TE-mediated duplicates, a similar research bias for retrogenes generated by LINE-1s also exists. With the exception of a couple of cases (*e.g.*, the aforementioned *SUN* locus [Xiao et al., 2008]), studies of duplicates mediated by LTR retrotransposons and DNA transposons are generally limited to transcriptomic or computational surveys rather than in-depth experimental characterization (Barbaglia et al., 2012; Tan et al., 2016, 2021; Zhao et al., 2018). Actually, we do not know the exact molecular function of *Ssk-FB4*, and we also lack a well-characterized chimeric protein emerging from the multiple gene fragments captured by Pack-MULEs or Pack-*Helitrons*. Since these genes are likely on the path of neofunctionalization, experimental characterization could lead to the discovery of another locus similar to *SUN* underlying important phenotypic renovation.

In conclusion, endowed with the intrinsic properties of TE-mediated duplications, the products tend to evolve novel functions through exon shuffling. Such pervasive phenomena highlight the mighty power of "controllers".

Conflict of interest

The authors declare no competing interests.

Acknowledgments

We are indebted to colleagues for their articles not covered in the review due to the limited space. We thank Dr. Jinfeng Chen for insightful comments. This work was supported by the Ministry of Agriculture and Rural Affairs of China, the National Key R&D Program of China (2019YFA0802600), the Chinese Academy of Sciences

(ZDBS-LY-SM005, XDPB17) and the National Natural Science Foundation of China (31970565).

References

Akiva, P., Toporik, A., Edelheit, S., Peretz, Y., Diber, A., Shemesh, R., Novik, A., Sorek, R., 2006. Transcription-mediated gene fusion in the human genome. Genome Res. 16, 30–36.

Almeida, M.V., Vernaz, G., Putman, A.L.K., Miska, E.A., 2022. Taming transposable elements in vertebrates: from epigenetic silencing to domestication. Trends Genet. 38, 529–553.

Babcock, M., Pavlicek, A., Spiteri, E., Kashork, C.D., Ioshikhes, I., Shaffer, L.G., Jurka, J., Morrow, B.E., 2003. Shuffling of genes within low-copy repeats on 22q11 (LCR22) by Alu-mediated recombination events during evolution. Genome Res. 13, 2519–2532.

Bailey, J.A., Liu, G., Eichler, E.E., 2003. An Alu transposition model for the origin and expansion of human segmental duplications. Am. J. Hum. Genet. 73, 823–834.

Barbaglia, A.M., Klusman, K.M., Higgins, J., Shaw, J.R., Hannah, L.C., Lal, S.K., 2012. Gene capture by Helitron transposons reshuffles the transcriptome of maize. Genetics 190, 965–975.

Batcher, K., Dickinson, P., Maciejczyk, K., Brzeski, K., Rasouliha, S.H., Letko, A., Droegemueller, C., Leeb, T., Bannasch, D., 2020. Multiple FGF4 retrocopies recently derived within canids. Genes 11, 839.

Batcher, K., Varney, S., York, D., Blacksmith, M., Kidd, J.M., Rebhun, R., Dickinson, P., Bannasch, D., 2022. Recent, full-length gene retrocopies are common in canids. Genome Res. 32, 1602–1611.

Betrán, E., Long, M., 2006. eLS Gene Fusion.

Brunner, S., Pea, G., Rafalski, A., 2005. Origins, genetic organization and transcription of a family of non-autonomous helitron elements in maize. Plant J. 43, 799–810.

Buzdin, A., Gogvadze, E., Kovalskaya, E., Volchkov, P., Ustyugova, S., Illarionova, A., Fushan, A., Vinogradova, T., Sverdlov, E., 2003. The human genome contains many types of chimeric retrogenes generated through in vivo RNA recombination. Nucleic Acids Res. 31, 4385–4390.

Buzdin, A., Ustyugova, S., Gogvadze, E., Vinogradova, T., Lebedev, Y., Sverdlov, E., 2002. A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3' terminus of L1. Genomics 80, 402–406.

Calatrava, V., Stephens, T.G., Gabr, A., Bhaya, D., Bhattacharya, D., Grossman, A.R., 2022. Retrotransposition facilitated the establishment of a primary plastid in the thecate amoeba Paulinella. Proc. Natl. Acad. Sci. U. S. A. 119, e2121241119.

Carelli, F.N., Hayakawa, T., Go, Y., Imai, H., Warnefors, M., Kaessmann, H., 2016. The life history of retrocopies illuminates the evolution of new mammalian genes. Genome Res. 26, 301–314.

Catoni, M., Jonesman, T., Cerruti, E., Paszkowski, J., 2019. Mobilization of Pack-CACTA transposons in Arabidopsis suggests the mechanism of gene shuffling. Nucleic Acids Res. 47, 1311–1320.

Cerbin, S., Jiang, N., 2018. Duplication of host genes by transposable elements. Curr. Opin. Genet. Dev. 49, 63–69.

Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal Jr., G., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J., et al., 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science 327, 302–305.

Ciomborowska, J., Rosikiewicz, W., Szklarczyk, D., Makalowski, W., Makalowska, I., 2013. "Orphan" retrogenes in the human genome. Mol. Biol. Evol. 30, 384–396.

Cosby, R.L., Chang, N.-C., Feschotte, C., 2019. Host-transposon interactions: conflict, cooperation, and cooption. Genes Dev. 33, 1098–1116.

Cosby, R.L., Judd, J., Zhang, R., Zhong, A., Garry, N., Pritham, E.J., Feschotte, C., 2021. Recurrent evolution of vertebrate transcription factors by transposase capture. Science 371, eabc6405.

Cost, G.J., Feng, Q.H., Jacquier, A., Boeke, J.D., 2002. Human L1 element target-primed reverse transcription in vitro. EMBO J. 21, 5899–5910.

Delviks-Frankenberry, K., Galli, A., Nikolaitchik, O., Mens, H., Pathak, V.K., Hu, W.-S., 2011. Mechanisms and factors that influence high frequency retroviral recombination. Viruses 3, 1650–1680.

Derr, L.K., Strathern, J.N., Garfinkel, D.J., 1991. RNA-mediated recombination in S. cerevisiae. Cell 67, 355–364.

Dobzhansky, T., 1973. Nothing in biology makes sense except in the light of evolution. Am. Biol. Teach. 35, 125–129.

Dong, Y., Lu, X., Song, W., Shi, L., Zhang, M., Zhao, H., Jiao, Y., Lai, J., 2011. Structural characterization of helitrons and their stepwise capturing of gene fragments in the maize genome. BMC Genomics 12, 1–11.

Dooner, H.K., Weil, C.F., 2007. Give-and-take: interactions between DNA transposons and their host plant genomes. Curr. Opin. Genet. Dev. 17, 486–492.

Dooner, H.K., Weil, C.F., 2013. Transposons and Gene Creation. Plant Transposons Genome Dynamics in Evolution, pp. 143–164.

Eickbush, T.H., 1992. Transposing without ends - the non-LTR retrotransposable elements. New Biol. 4, 430–440.

Elrouby, N., Bureau, T.E., 2010. Bs1, a new chimeric gene formed by retrotransposon-mediated exon shuffling in maize. Plant Physiol. 153, 1413–1424.

Emerson, J.J., Cardoso-Moreira, M., Borevitz, J.O., Long, M., 2008. Natural selection shapes genome-wide patterns of copy-number polymorphism in Drosophila melanogaster. Science 320, 1629–1631.

Emerson, J.J., Kaessmann, H., Betran, E., Long, M.Y., 2004. Extensive gene traffic on the mammalian X chromosome. Science 303, 537–540.

Fanning, T.G., Singer, M.F., 1987. LINE-1 - a mammalian transposable element. Biochim. Biophys. Acta 910, 203–212.

Feng, Q.H., Moran, J.V., Kazazian, H.H., Boeke, J.D., 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87, 905–916.

Feschotte, C., Wessler, S.R., 2001. Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes. Proc. Natl. Acad. Sci. U. S. A. 98, 8923–8924.

Finnegan, D.J., 1989. Eukaryotic transposable elements and genome evolution. Trends Genet. 5, 103–107.

Fueyo, R., Judd, J., Feschotte, C., Wysocka, J., 2022. Roles of transposable elements in the regulation of mammalian transcription. Nat. Rev. Mol. Cell Biol. 23, 481–497.

Gilbert, W., 1978. Why genes in pieces. Nature 271, 501.

Gloor, G.B., Nassif, N.A., Johnsonschlitz, D.M., Preston, C.R., Engels, W.R., 1991. Targeted gene replacement in Drosophila via P-element-induced gap repair. Science 253, 1110–1117.

Goodier, J.L., Ostertag, E.M., Kazazian, H.H., 2000. Transduction of 3'-flanking sequences is common in L1 retrotransposition. Hum. Mol. Genet. 9, 653–657.

Goodrich, D.W., Duesberg, P.H., 1990. Retroviral recombination during reverse transcription. Proc. Natl. Acad. Sci. U. S. A. 87, 2052–2056.

Grabundzija, I., Hickman, A.B., Dyda, F., 2018. Helraiser intermediates provide insight into the mechanism of eukaryotic replicative transposition. Nat. Commun. 9, 1278.

Grabundzija, I., Messing, S.A., Thomas, J., Cosby, R.L., Bilic, I., Miskey, C., Gogol-Doering, A., Kapitonov, V., Diem, T., Dalda, A., et al., 2016. A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. Nat. Commun. 7, 10716.

Gray, Y.H.M., 2000. It takes two transposons to tango - transposable-element-mediated chromosomal rearrangements. Trends Genet. 16, 461–468.

Hajjar, A.M., Linial, M.L., 1993. A model system for nonhomologous recombination between retroviral and cellular RNA. J. Virol. 67, 3845–3853.

Han, M.-J., Shen, Y.-H., Xu, M.-S., Liang, H.-Y., Zhang, H.-H., Zhang, Z., 2013. Identification and evolution of the silkworm Helitrons and their contribution to transcripts. DNA Res. 20, 471–484.

Hanada, K., Vallejo, V., Nobuta, K., Slotkin, R.K., Lisch, D., Meyers, B.C., Shiu, S.-H., Jiang, N., 2009. The functional role of pack-MULEs in rice inferred from purifying selection and expression profile. Plant Cell 21, 25–38.

Hu, Y., Wu, X., Jin, G., Peng, J., Leng, R., Li, L., Gui, D., Fan, C., Zhang, C., 2022. Rapid genome evolution and adaptation of Thlaspi arvense mediated by recurrent RNA-based and tandem gene duplications. Front. Plant Sci. 12, 772655.

Hwang, S.-Y., Jung, H., Mun, S., Lee, S., Park, K., Baek, S.C., Moon, H.C., Kim, H., Kim, B., Choi, Y., et al., 2021. L1 retrotransposons exploit RNA m(6)A modification as an evolutionary driving force. Nat. Commun. 12, 880.

Innan, H., Kondrashov, F., 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat. Rev. Genet. 11, 97–108.

Izsvak, Z., Stuwe, E., Fiedler, D., Katzer, A., Jeggo, P., Ivics, Z., 2004. Healing the wounds inflicted by Sleeping Beauty transposition by double-strand break repair in mammalian somatic cells. Eur. J. Cell Biol. 83, 279–290.

Jamain, S., Girondot, M., Leroy, P., Clergue, M., Quach, H., Fellous, M., Bourgeron, T., 2001. Transduction of the human gene *FAM8A1* by endogenous retrovirus during primate evolution. Genomics 78, 38–45.

Jiang, N., Bao, Z.R., Zhang, X.Y., Eddy, S.R., Wessler, S.R., 2004. Pack-MULE transposable elements mediate gene evolution in plants. Nature 431, 569–573.

Jiang, N., Gao, D., Xiao, H., van der Knaap, E., 2009. Genome organization of the tomato sun locus and characterization of the unusual retrotransposon Rider. Plant J. 60, 181–193.

Kaessmann, H., 2009. Genetics. More than just a copy. Science 325, 958–959.

Kaessmann, H., Vinckenbosch, N., Long, M., 2009. RNA-based gene duplication: mechanistic and evolutionary insights. Nat. Rev. Genet. 10, 19–31.

Kapitonov, V.V., Jurka, J., 2007. Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet. 23, 521–529.

Katju, V., 2012. In with the old, in with the new: the promiscuity of the duplication process engenders diverse pathways for novel gene creation. Int. J. Evol. Biol. 2012, 341932.

Kazazian, H.H., 2004. Mobile elements: drivers of genome evolution. Science 303, 1626–1632.

Kazazian, H.H., Moran, J.V., 1998. The impact of L1 retrotransposons on the human genome. Nat. Genet. 19, 19–24.

Kleckner, N., Chan, R.K., Tye, B.-K., Botstein, D., 1975. Mutagenesis by insertion of a drug-resistance element carrying an inverted repetition. J. Mol. Biol. 97, 561–575.

Kojima, K.K., Bao, W., Kojima, N.F., Kohany, O., 2023. Repbase 2022 Year in Review.

Kosek, D., Grabundzija, I., Lei, H., Bilic, I., Wang, H., Jin, Y., Peaslee, G.F., Hickman, A.B., Dyda, F., 2021. The large bat Helitron DNA transposase forms a compact monomeric assembly that buries and protects its covalently bound 5'-transposon end. Mol. Cell 81, 4271–4286.

Kubiak, M.R., Makalowska, I., 2017. Protein-coding genes' retrocopies and their functions. Viruses 9, 80.

Lal, S., Oetjens, M., Hannah, L.C., 2009. Helitrons: enigmatic abductors and mobilizers of host genome sequences. Plant Sci. (Amsterdam, Neth.) 176, 181–186.

Lisch, D., 2005. Pack-MULEs: theft on a massive scale. Bioessays 27, 353–355.

Long, M., Betran, E., Thornton, K., Wang, W., 2003. The origin of new genes: glimpses from the young and old. Nat. Rev. Genet. 4, 865–875.

Long, M., VanKuren, N.W., Chen, S., Vibranovski, M.D., 2013. New gene evolution: little did we know. Annu. Rev. Genet. 47, 307–333.
Long, M.Y., Langley, C.H., 1993. Natural-selection and the origin of jingwei, a chimeric processed functional gene in Drosophila. Science 260, 91–95.
Luan, D.D., Korman, M.H., Jakubczak, J.L., Eickbush, T.H., 1993. Reverse transcription of R2BM RNA is primed by a nick at the chromosomal target site - a mechanism for non-LTR retrotransposition. Cell 72, 595–605.
Lynch, M., Conery, J.S., 2000. The evolutionary fate and consequences of duplicate genes. Science 290, 1151–1155.
Lynch, M., Katju, V., 2004. The altered evolutionary trajectories of gene duplicates. Trends Genet. 20, 544–549.
Ma, C., Li, C., Ma, H., Yu, D., Zhang, Y., Zhang, D., Su, T., Wu, J., Wang, X., Zhang, L., et al., 2022. Pan-cancer surveys indicate cell cycle-related roles of primate-specific genes in tumors and embryonic cerebrum. Genome Biol. 23, 1–29.
Makalowski, W., Mitchell, G.A., Labuda, D., 1994. Alu sequences in the coding regions of messenger-RNA - source of protein variability. Trends Genet. 10, 188–193.
Martin-Alonso, S., Frutos-Beltran, E., Menendez-Arias, L., 2021. Reverse transcriptase: from transcriptomics to genome editing. Trends Biotechnol. 39, 194–210.
McClintock, B., 1948. Mutable loci in maize. Carnegie Inst. Wash. 47, 155–169.
McClintock, B., 1950. The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. U. S. A. 36, 344–355.
McClintock, B., 1956. Controlling elements and the gene. Cold Spring Harb. Symp. Quant. Biol. 21, 197–216.
Moran, J.V., DeBerardinis, R.J., Kazazian, H.H., 1999. Exon shuffling by L1 retrotransposition. Science 283, 1530–1534.
Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., Kazazian, H.H., 1996. High frequency retrotransposition in cultured mammalian cells. Cell 87, 917–927.
Moran, J.V., Zimmerly, S., Eskes, R., Kennell, J.C., Lambowitz, A.M., Butow, R.A., Perlman, P.S., 1995. Mobile group-II introns of yeast mitochondrial-DNA are novel site-specific retroelements. Mol. Cell. Biol. 15, 2828–2838.
Morgante, M., Brunner, S., Pea, G., Fengler, K., Zuccolo, A., Rafalski, A., 2005. Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat. Genet. 37, 997–1002.
Nei, M., 2007. The new mutation theory of phenotypic evolution. Proc. Natl. Acad. Sci. U. S. A. 104, 12235–12242.
Ohno, S., 1970. Evolution by Gene Duplication. London: George Alien & Unwin Ltd. Berlin, Heidelberg and New York: Springer-Verlag.
Pan, D., Zhang, L., 2009. Burst of young retrogenes and independent retrogene formation in mammals. PLoS One 4, e5040.
Parker, H.G., VonHoldt, B.M., Quignon, P., Margulies, E.H., Shao, S., Mosher, D.S., Spady, T.C., Elkahloun, A., Cargill, M., Jones, P.G., et al., 2009. An expressed *Fgf4* retrogene is associated with breed-defining chondrodysplasia in domestic dogs. Science 325, 995–998.
Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., et al., 2009. The Sorghum bicolor genome and the diversification of grasses. Nature 457, 551–556.
Peng, H., Mirouze, M., Bucher, E., 2022. Extrachromosomal circular DNA: a neglected nucleic acid molecule in plants. Curr. Opin. Plant Biol. 69, 102263.
Petrov, D.A., Lozovskaya, E.R., Hartl, D.L., 1996. High intrinsic: rate of DNA loss in Drosophila. Nature 384, 346–349.
Pickeral, O.K., Makalowski, W., Boguski, M.S., Boeke, J.D., 2000. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. Genome Res. 10, 411–415.
Pritham, E.J., Feschotte, C., 2007. Massive amplification of rolling-circle transposons in the lineage of the bat Myotis lucifugus. Proc. Natl. Acad. Sci. U. S. A. 104, 1895–1900.
Qian, W., Zhang, J., 2008. Gene dosage and gene duplicability. Genetics 179, 2319–2324.
Rosenberg, S.M., Queitsch, C., 2014. Combating evolution to fight disease. Science 343, 1088–1089.
Rosikiewicz, W., Kabza, M., Kosinski, J.G., Ciomborowska-Basheer, J., Kubiak, M.R., Makalowska, I., 2017. RetrogeneDB-a database of plant and animal retrocopies. Database 2017, bax038.
Rubin, G.M., Kidwell, M.G., Bingham, P.M., 1982. The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell 29, 987–994.
Sayah, D.M., Sokolskaja, E., Berthoux, L., Luban, J., 2004. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. Nature 430, 569–573.
Schacherer, J., Tourrette, Y., Souciet, J.L., Potier, S., De Montigny, J., 2004. Recovery of a function involving gene duplication by retroposition in *Saccharomyces cerevisiae*. Genome Res. 14, 1291–1297.
Sulak, M., Fong, L., Mika, K., Chigurupati, S., Yon, L., Mongan, N.P., Emes, R.D., Lynch, V.J., 2016. TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. Elife 5, e11994.
Sun, C., Shepard, D.B., Chong, R.A., Arriaza, J.L., Hall, K., Castoe, T.A., Feschotte, C., Pollock, D.D., Mueller, R.L., 2012. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. Genome Biol. Evol. 4, 168–183.
Tan, S., Cardoso-Moreira, M., Shi, W., Zhang, D., Huang, J., Mao, Y., Jia, H., Zhang, Y., Chen, C., Shao, Y., et al., 2016. LTR-mediated retroposition as a mechanism of RNA-based duplication in metazoans. Genome Res. 26, 1663–1675.
Tan, S., Ma, H., Wang, J., Wang, M., Wang, M., Yin, H., Zhang, Y., Zhang, X., Shen, J., Wang, D., et al., 2021. DNA transposons mediate duplications via transposition-independent and -dependent mechanisms in metazoans. Nat. Commun. 12, 4280.
Tempel, S., Nicolas, J., El Amrani, A., Couee, I., 2007. Model-based identification of Helitrons results in a new classification of their families in *Arabidopsis thaliana*. Gene 403, 18–28.
Thomas, J., Phillips, C.D., Baker, R.J., Pritham, E.J., 2014. Rolling-circle transposons catalyze genomic innovation in a mammalian lineage. Genome Biol. Evol. 6, 2595–2610.
Thomas, J., Pritham, E.J., 2015. Helitrons, the eukaryotic rolling-circle transposable elements. In: Mobile DNA iii, pp. 891–924.
Tsubota, S.I., Huong, D.V., 1991. Capture of flanking DNA by a P-element in Drosophila-melanogaster - creation of a transposable element. Proc. Natl. Acad. Sci. U. S. A. 88, 693–697.
Vinckenbosch, N., Dupanloup, I., Kaessmann, H., 2006. Evolutionary fate of retroposed gene copies in the human genome. Proc. Natl. Acad. Sci. U. S. A. 103, 3220–3225.
Wang, D., Yu, C., Zuo, T., Zhang, J., Weber, D.F., Peterson, T., 2015. Alternative transposition generates new chimeric genes and segmental duplications at the maize p1 locus. Genetics 201, 925–935.
Wang, L., Tracy, L., Su, W., Yang, F., Feng, Y., Silverman, N., Zhang, Z.Z.Z., 2022a. Retrotransposon activation during Drosophila metamorphosis conditions adult antiviral responses. Nat. Genet. 1–13.
Wang, W., Brunet, F.G., Nevo, E., Long, M., 2002. Origin of sphinx, a young chimeric RNA gene in Drosophila melanogaster. Proc. Natl. Acad. Sci. U. S. A. 99, 4448–4453.
Wang, W., Zheng, H., Fan, C., Li, J., Shi, J., Cai, Z., Zhang, G., Liu, D., Zhang, J., Vang, S., et al., 2006. High rate of chimeric gene origination by retroposition in plant genomes. Plant Cell 18, 1791–1802.
Wang, X., Yan, X., Hu, Y., Qin, L., Wang, D., Jia, J., Jiao, Y., 2022b. A recent burst of gene duplications in Triticeae. Plant Commun. 3, 100286.
Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D., Moran, J.V., 2001. Human L1 retrotransposition: cis preference versus trans complementation. Mol. Cell. Biol. 21, 1429–1439.
Wells, J.N., Feschotte, C., 2020. A field guide to eukaryotic transposable elements. Annu. Rev. Genet. 54, 539–561.
Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al., 2007. A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8, 973–982.
Wilhelm, M., Wilhelm, F.X., 2001. Reverse transcription of retroviruses and LTR retrotransposons. Cell. Mol. Life Sci. 58, 1246–1262.
Witt, E., Benjamin, S., Svetec, N., Zhao, L., 2019. Testis single-cell RNA-seq reveals the dynamics of de novo gene transcription and germline mutational bias in Drosophila. Elife 8, e47138.
Xia, S., Wang, Z., Zhang, H., Hu, K., Zhang, Z., Qin, M., Dun, X., Yi, B., Wen, J., Ma, C., et al., 2016. Altered transcription and neofunctionalization of duplicated genes rescue the harmful effects of a chimeric gene in Brassica napus. Plant Cell 28, 2060–2078.
Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J., van der Knaap, E., 2008. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science 319, 1527–1530.
Xie, K.T., Wang, G., Thompson, A.C., Wucherpfennig, J.I., Reimchen, T.E., MacColl, A.D.C., Schluter, D., Bell, M.A., Vasquez, K.M., Kingsley, D.M., 2019. DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. Science 363, 81–84.
Yang, L., Emerman, M., Malik, H.S., McLaughlin, R.N.J.E., 2020. Retrocopying expands the functional repertoire of APOBEC3 antiviral proteins in primates. Elife 9, e58436.
Yang, L.X., Bennetzen, J.L., 2009. Distribution, diversity, evolution, and survival of Helitrons in the maize genome. Proc. Natl. Acad. Sci. U. S. A. 106, 19922–19927.
Yang, S., Arguello, J.R., Li, X., Ding, Y., Zhou, Q., Chen, Y., Zhang, Y., Zhao, R., Brunet, F., Peng, L., et al., 2008. Repetitive element-mediated recombination as a mechanism for new gene origination in Drosophila. PLoS Genet. 4, e3.
Yant, S.R., Kay, M.A., 2003. Nonhomologous-end-joining factors regulate DNA repair fidelity during Sleeping Beauty element transposition in mammalian cells. Mol. Cell. Biol. 23, 8505–8518.
Zhang, D., Leng, L., Chen, C., Huang, J., Zhang, Y., Yuan, H., Ma, C., Chen, H., Zhang, Y.E., 2022. Dosage sensitivity and exon shuffling shape the landscape of polymorphic duplicates in Drosophila and humans. Nat. Ecol. Evol. 6, 273–287.
Zhang, J., Yang, H., Long, M., Li, L., Dean, A.M., 2010. Evolution of enzymatic activities of testis-specific short-chain dehydrogenase/reductase in Drosophila. J. Mol. Evol. 71, 241–249.
Zhang, J.M., Dean, A.M., Brunet, F., Long, M.Y., 2004a. Evolving protein functional diversity in new genes of Drosophila. Proc. Natl. Acad. Sci. U. S. A. 101, 16246–16250.
Zhang, J.Z., 2003. Evolution by gene duplication: an update. Trends Ecol. Evol. 18, 292–298.

Zhang, W., Tautz, D., 2022. Tracing the origin and evolutionary fate of recent gene retrocopies in natural populations of the house mouse. Mol. Biol. Evol. 39, msab360.
Zhang, Y., Lu, S., Zhao, S., Zheng, X., Long, M., Wei, L., 2009. Positive selection for the male functionality of a co-retroposed gene in the hominoids. BMC Evol. Biol. 9, 1–12.
Zhang, Z.L., Carriero, N., Gerstein, M., 2004b. Comparative analysis of processed pseudogenes in the mouse and human genomes. Trends Genet. 20, 62–67.
Zhang, Z.L., Harrison, P.M., Liu, Y., Gerstein, M., 2003. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. Genome Res. 13, 2541–2558.
Zhao, D., Ferguson, A.A., Jiang, N., 2016. What makes up plant genomes: the vanishing line between transposable elements and genes. Biochim. Biophys. Acta - Gene Regul. Mech. 1859, 366–380.
Zhao, D., Hamilton, J.P., Vaillancourt, B., Zhang, W., Eizenga, G.C., Cui, Y., Jiang, J., Buell, C.R., Jiang, N., 2018. The unique epigenetic features of Pack-MULEs and their impact on chromosomal base composition and expression spectrum. Nucleic Acids Res. 46, 2380–2397.
Zhou, Y., Zhang, C., 2019. Evolutionary patterns of chimeric retrogenes in *Oryza* species. Sci. Rep. 9, 17733.