Repetitive Sequences in Complex Genomes: Structure and Evolution

Jerzy Jurka, Vladimir V. Kapitonov, Oleksiy Kohany, and Michael V. Jurka

Genetic Information Research Institute, Mountain View, California 94043; email: jurka@girinst.org, vladimir@girinst.org, kohany@girinst.org, michael@girinst.org

Annu. Rev. Genomics Hum. Genet. 2007. 8:241–59

First published online as a Review in Advance on May 21, 2007.

The *Annual Review of Genomics and Human Genetics* is online at genom.annualreviews.org

This article's doi: 10.1146/annurev.genom.8.080706.092416

1527-8204/07/0922-0241$20.00

Key Words

transposable elements, repetitive DNA, regulation, speciation

Abstract

Eukaryotic genomes contain vast amounts of repetitive DNA derived from transposable elements (TEs). Large-scale sequencing of these genomes has produced an unprecedented wealth of information about the origin, diversity, and genomic impact of what was once thought to be "junk DNA." This has also led to the identification of two new classes of DNA transposons, *Helitrons* and *Polintons*, as well as several new superfamilies and thousands of new families. TEs are evolutionary precursors of many genes, including RAG1, which plays a role in the vertebrate immune system. They are also the driving force in the evolution of epigenetic regulation and have a long-term impact on genomic stability and evolution. Remnants of TEs appear to be overrepresented in transcription regulatory modules and other regions conserved among distantly related species, which may have implications for our understanding of their impact on speciation.

INTRODUCTION

The term "repetitive sequences" (repeats, DNA repeats, repetitive DNA) refers to homologous DNA fragments that are present in multiple copies in the genome. Repetitive DNA was originally discovered based on reassociation kinetics and classified into "highly" and "middle" repetitive sequences (14), roughly corresponding to tandem and interspersed repeats discussed below. This review is centered primarily on repeat research based on DNA sequence analysis and does not cover the so-called low copy repeats (LCRs), also known as segmental duplications, which represent a separate category of duplicated diverse chromosomal segments (105).

Repeats can be clustered into distinct families each traceable to a single ancestral sequence or a closely related group of ancestral sequences. In contrast to multigene families, which are defined based on their biological role, repetitive families are usually defined based on their active ancestors, called master or source genes, and on their generation mechanisms. Over time, individual elements from repetitive families may acquire diverse biological roles.

There are two basic types of repetitive sequences: interspersed repeats and tandem repeats. Interspersed repeats are DNA fragments with an upper size limit of 20–30 kb, inserted more or less at random into host DNA. In contrast, tandem repeats represent arrays of DNA fragments immediately adjacent to each other in head-to-tail orientation. This review focuses on interspersed repetitive DNA from eukaryotic genomes. Interspersed repeats are mostly inactive and often incomplete copies of transposable elements (TEs) inserted into genomic DNA. TEs are segments of DNA or RNA capable of being reproduced and inserted in the host genome. At the same time, genomes are essentially conservative structures that have evolved mechanisms to counteract such insertions. Therefore, TEs and host genomes are locked in a permanent antagonistic relationship resembling an "arms race." Eukaryotic hosts continuously suppress activities of TEs, but TE proliferation persists in virtually all known eukaryotic species. Of all eukaryotic genomes sequenced to date, only the genome *Plasmodium falciparum* appears not to host any active TEs (35).

Why do complex, conservative genomes tolerate the activities of inherently antagonistic elements? TEs cannot be easily eliminated and their endurance in the host can be compared to that of parasites. Furthermore, if TEs can provide evolutionary advantages to the host, their chances of survival increase. The view that TEs are beneficial to the host is not new (16, 44, 68, 87) but recent progress in the field puts it squarely at the center of the ongoing debate on eukaryotic evolution.

STRUCTURE AND SYSTEMATICS OF TRANSPOSABLE ELEMENTS

General Characteristics

Figure 1 presents a schematic structure of TEs. All types of TEs are represented by autonomous and nonautonomous variants. Whereas an autonomous element encodes a complete set of enzymes characteristic of its family and is self-sufficient in terms of transposition, a nonautonomous element transposes by borrowing the protein machinery encoded by its autonomous relatives. Despite their dazzling diversity, all eukaryotic TEs fall into two basic types: retrotransposons and DNA transposons. Retrotransposons are transposed through an RNA intermediate. Their messenger RNA (mRNA) is expressed in the host cell, reverse transcribed, and the resulting complementary DNA (cDNA) copy is integrated back into the host genome. Reverse transcription and integration are catalyzed by reverse transcriptase (RT) and endonuclease/integrase (EN/INT), which are encoded by autonomous elements. Unlike retrotransposons, DNA transposons

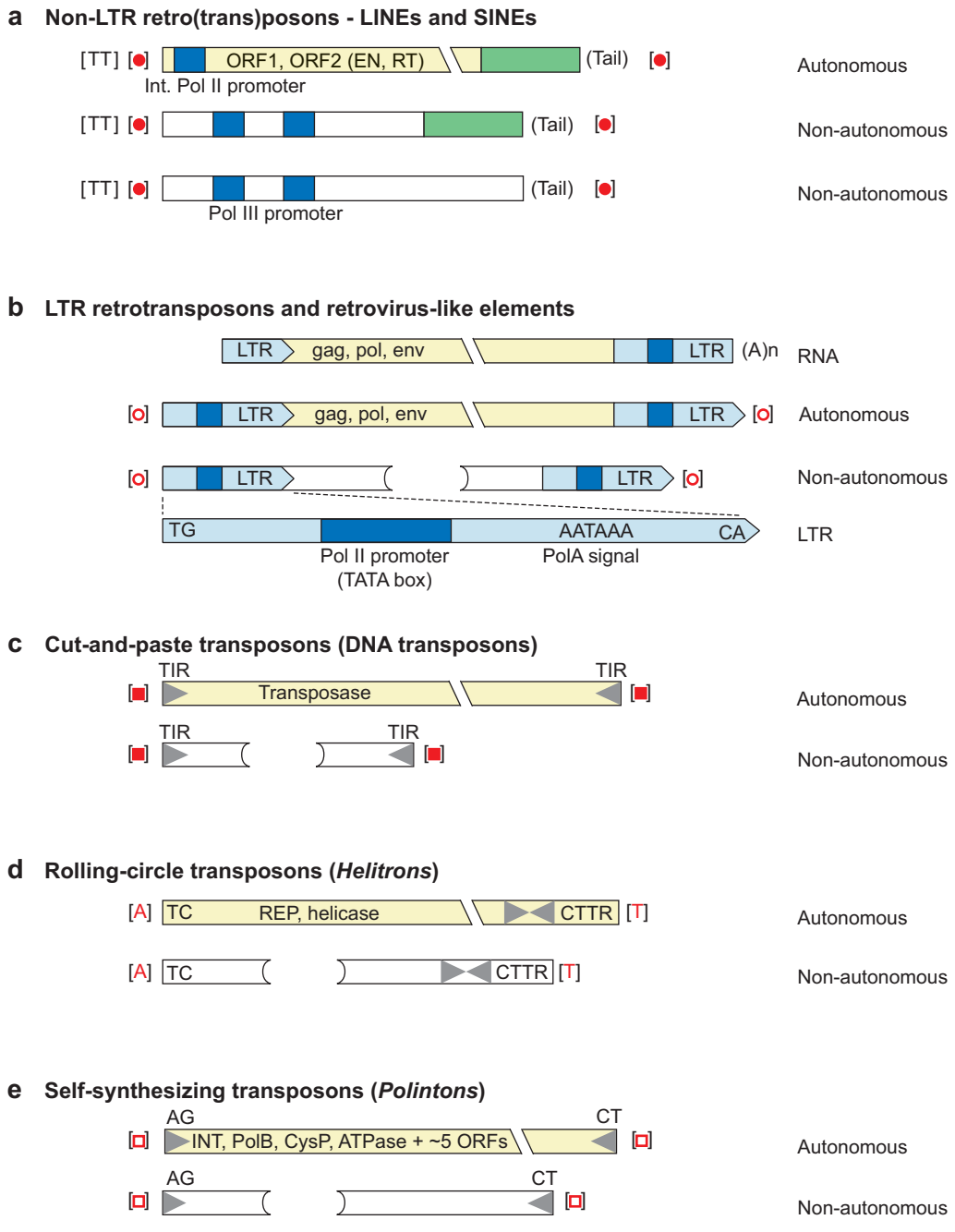

Figure 1

A schematic representation of major classes of transposable elements, including nonautonomous elements.

are transposed by moving their genomic DNA copies from one chromosomal location to another without any RNA intermediate. Most retrotransposons and DNA transposons are flanked by target site duplications (TSDs) resulting from fill-in repair of staggered nicks generated at the DNA target site upon insertion of TEs (42).

All currently known eukaryotic retrotransposons can be divided into four classes: non-long terminal repeat (LTR) retrotransposons, LTR retrotransposons, *Penelope*, and *DIRS* retrotransposons. Although the first two classes (**Figure 1*a*,*b***) are relatively well established and studied (29), the *Penelope* and *DIRS* classes were only recently introduced (2, 30, 81, 98). Members of all four classes of retrotransposons are present in the genomes of virtually all eukaryotic kingdoms: Protista, Plantae, Fungi, and Animalia. The only exception is *Penelope*, which, so far, has not been identified in plants.

Eukaryotic DNA transposons belong to three classes: "cut-and-paste" transposons, *Helitrons*, and *Polintons* (**Figure 1*c*,*d*,*e***). The corresponding mechanisms of transposition are cut-and-paste (23), rolling-circle replicative (60), and self-synthesizing (65), respectively. The cut-and-paste transposons and *Helitrons* cannot synthesize their own DNA; instead, they multiply using host replication machinery.

Non-Long Terminal Repeat and Long Terminal Repeat Retrotransposons

A typical autonomous non-LTR retrotransposon, commonly referred to as a long interspersed element (LINE), contains one or two open reading frames (ORFs). It includes an internal promoter in the 5′ terminal region that governs transcription of the retrotransposon inserted in the host genome. The mechanism of LINE retrotransposition and integration into the genome is well studied and is viewed as a coupled process called target-primed reverse transcription (TPRT). According to the TPRT model, reverse transcription is primed by the free 3′ hydroxyl group at the target DNA nick introduced by EN (29). The model was recently enhanced by the finding that initiation of the *L1* reverse transcription does not require base pairing between the primer and template (72). Moreover, as expected from the model, EN is not necessary for *L1* retrotransposition when free 3′-hydroxyl groups become available in disfunctional telomeres (91). Both RT and EN domains in *L1* are encoded by the same ORF. An mRNA expressed during transcription of a genomic copy of LINE retrotransposon serves as a template for reverse transcription, and the resulting cDNA is inserted in the genome.

Based on structural features of non-LTR retrotransposons and phylogeny of RTs, LINEs can be assigned to five groups, called *R2*, *L1*, *RTE*, *I*, and *Jockey*, which can be subdivided into 15 clades (29, 70). It is believed that the *R2* group is composed of the most ancient non-LTR retrotransposons, the *CRE*, *NeSL*, *R2*, and *R4* clades, which are characterized by a single ORF coding for RT and an EN C terminal to the RT domain. The *R2* EN is similar to different restriction enzymes, and all TEs from the *R2* group retrotranspose into highly specific target sites. Members of the remaining four groups encode the apurinic-apyrimidinic endonuclease (APE), which is always N terminal to the RT domain. In addition to RT and EN, members of the first group code for RNase H (29), including the *Ingi*, *I*, *LOA*, *R1*, and *Tad1* clades. Nonautonomous non-LTR retrotransposons are usually referred to as short interspersed elements (SINE) retrotransposons. Typically, they are mosaic structures derived from transfer RNA (tRNA) or 7SL or 5S ribosomal RNA, and contain 5′ internal pol III promoters involved in transcription. The 3′ ends of SINEs are either derived from LINE elements or contain poly(A) tails recognizable by *L1* elements. They may share common structural constraints (111). Their retrotransposition is catalyzed by RT/EN encoded by the autonomous non-LTR retrotransposons.

Non-LTR retrotransposons are transmitted vertically (i.e., from parents to offspring), with some notable exceptions (71).

An LTR retrotransposon (**Figure 1*b***) may carry three ORFs coding for the *gag*, *env*, and *pol* proteins. The *pol* protein is composed of the RT, EN, and aspartyl protease domains. The EN domain in LTR retrotransposons is usually called INT and is distantly related to the DDE transposase (named after two aspartate and one glutamate residues forming a catalytic triad), encoded by *Mariner* DNA transposons (20, 29). LTR retrotransposons can be transferred horizontally (49), although the extent of the process is not clear.

Penelope. The *Penelope* retrotransposons encode a single ORF composed of the RT and EN domains. The latter is similar to GIY-YIG intron-encoded ENs (2, 30, 84, 121), named after the conserved amino acid motif Gly-Ile-Tyr-Xn-Tyr-Ile-Gly. It appears that the *Penelope* RT is closer to telomerases and bacterial RTs than RTs encoded by non-LTR retrotransposons (2). Like many families of non-LTR retrotransposons, *Penelope* elements generate 10–15 base pair (bp) TSDs and probably follow the TPRT model of retrotransposition (29, 30). However, *Penelope* elements are characterized by unusual LTR-like or inverted terminal repeats not typical for standard non-LTR retrotransposons. Also, some *Penelope* elements from different species retain introns after their retrotransposition (2). Based on their structural and phylogenetic features, *Penelopes* are viewed as a separate class of retrotransposons. However, given the low resolution of phylogenetic trees built for extremely divergent and ancient RTs, an alternative view of *Penelopes* as the most ancient/basal group of non-LTR retrotransposons cannot yet be ruled out.

DIRS. The *Dictyostelium* intermediate repeat sequence (*DIRS*) retrotransposons encode a RT that is phylogenetically closer to that encoded by LTR retrotransposons than to the RT in non-LTR retrotransposons (29). Until recently, *DIRS* elements were viewed as an enigmatic class of INT-free retrotransposons characterized by an unusual structure of terminal repeats (19, 29). However, it turns out that *DIRS* elements encode a protein belonging to the INT family of tyrosine recombinases (tyrosine INT) (40). This observation and the unusual structure of termini led to the classification of *DIRS* elements as a separate class of LTR retrotransposons (28, 41). Given a wide distribution of highly diverse *DIRS* retrotransposons in different eukaryotic kingdoms, it appears that they are as ancient as LTR retrotransposons. It also appears that the *DIRS* RT is grouped phylogenetically with the *Gypsy* RT (29, 98) and separately from the *BEL* and *Copia* LTR retrotransposons, which are viewed as the most ancient LTR retrotransposons (29). Therefore, the most parsimonious scenario of *DIRS* origin is that they evolved from a *Gypsy*-like ancestral LTR retrotransposon after recruiting the tyrosine INT, which replaced the standard DDE INT. This scenario is consistent with the observation of tyrosine recombinase/INT-encoding DNA transposon-like elements in some fungi (*Crypton* transposons) (39) and in ciliates (*Tec* transposons) (27, 48). The suggested recruitment might have occurred following insertion of an ancient tyrosine recombinase-encoding DNA transposon into the *Gypsy*-like predecessor of *DIRS* elements. Analogously to *Penelope* retrotransposons, some *DIRS* elements retain introns in ORFs (41). Such intron retention could be important for the retrotranspositon of *Penelope* and *DIRS*. For instance, non-spliced *DIRS/Penelope* mRNA retained in the nucleus can be a better substrate for retrotransposition than the spliced one (30, 41).

Cut-and-Paste DNA Transposons

During its transposition, a cut-and-paste DNA transposon is excised (cut) from its original genomic location and inserted (pasted) into a new site (23). Both reactions are catalyzed by a transposase that binds the termini of a transposon and its target site and

Table 1 Superfamilies of "cut-and-paste" DNA transposons

Superfamily name	Related bacterial transposases	Size of target site duplications
En/Spm		3
hAT		8
Harbinger	IS5	3
IS4EU	IS4	2
Mariner	IS630	2
Merlin	IS1016	8–9
Mirage		2
MuDR	IS256	9–10
Novosib		8
P		7–8
piggyBac		4
Rehavkus		9
Transib		5

introduces DNA nicks. Most DNA transposons contain 10–400 bp long terminal inverted repeats (TIRs) at both ends. However, in some active transposons, TIRs are imperfect or absent (e.g., some *MuDR* transposons in *Arabidopsis thaliana*) (59). Eukaryotic "cut-and-paste" DNA transposons can be assigned to 13 superfamilies (**Table 1**). Each superfamily includes diverse families composed of autonomous and nonautonomous elements, whose transposition is catalyzed by superfamily-specific transposases. Transposases from different superfamilies are not similar to each other [i.e., position specific iterative basic local alignment and search tool (PSI-BLAST) expected values are higher than 0.05] (1). In addition to the superfamily-specific transposases, each superfamily is characterized by a specific length of TSDs (**Table 1**). However, some superfamilies, such as *En/Spm* and *Harbinger*, have the same length of TSDs. Autonomous DNA transposons from most superfamilies encode only one protein (transposase). They include *Mariner* (97), *hAT* (73), *P* (4), *piggyBac* (85), *Transib* (61, 64), *Merlin* (31), *Mirage*, *IS4EU*, *Novosib* and *Rehavkus* (36). Transposons from the *En/Spm* (73), *Harbinger* (59), and *MuDR* (122) superfamilies code for DNA-binding proteins in addition to transposases.

Harbinger was the first superfamily of DNA transposons discovered based on computational studies (59). The autonomous *Harbingers* encode two proteins: a ~400-amino acid (aa) *Harbinger* transposase and a ~200-aa DNA-binding protein that includes the conserved SANT/myb/trihelix motif. The *Harbinger* transposase is distantly related to transposases encoded by the IS5 group of bacterial transposons, including IS5, IS112, and ISL2. *Harbingers* are typically flanked by 3-bp TSDs, frequently TAA or TTA trinucleotides, but some *Harbingers* from the zebrafish genome show a striking preference for a 17-bp target site (AAAACACCWG-GTCTTTT), longer than the target for any other DNA transposon family (62).

Helitrons. *Helitron* DNA transposons transpose via replicative rolling-circle transposition (60). *Helitrons* are present in the genomes of plants, fungi, insects, nematodes, and vertebrates. In some species, including *A. thaliana* and *Caenorhabditis elegans*, they constitute ~2% of the genome. Autonomous *Helitrons* encode the ~1500-aa, so-called Rep/Hel protein, composed of the replication initiator (Rep), and helicase (Hel) conserved domains. The Rep domain spans a ~160-aa region composed of the "two-His" (E-FYW-Q-K-R-G-LAV-PVH-X-H) and "KYK" (Yg-LVW-FAT-Kq-Y-X-X-K) motifs separated by ~130 aa. These motifs are conserved in Reps, which are encoded by plasmids and single-stranded DNA viruses replicating by rolling-circle mechanism. The Rep proteins perform both cleavage and ligation of DNA during rolling-circle replication, the same as transposases. The ~500-aa Hel domain is a helicase that belongs to the SF1 superfamily of DNA helicases. *Helitron* is the only known class of transposons in eukaryotes that integrates into the genome without introducing TSDs. Usually, the *Helitron* integration occurs precisely between A and T nucleotides in the host. *Helitrons* do not have TIRs, which are typically present in other DNA transposons. Instead, *Helitrons* have conserved 5′-TC and CTRR-3′

termini. They also contain a ~18-bp hairpin separated by 10–12 nucleotides from the 3′ end. Presumably, the hairpin serves as the terminator of rolling-circle replication, which is believed to be the mechanism for *Helitron*'s transposition. So far, only *Helitrons* found in the *Aspergillus nidulans* genome do not contain the 3′ hairpin (33).

Although no active *Helitrons* have been isolated and studied experimentally so far, the main features of *Helitron* transposition can be predicted a priori based on the structural invariants detected in different *Helitrons* and known properties of bacterial rolling-circle replicons (60). *Helitron* transposition starts from a site-specific Rep-encoded nicking of the transposon-plus strand. Next, the free 3′-OH end of the nicked-plus strand serves as a primer for leading-strand DNA synthesis facilitated by the *Helitron* helicase and some host replication proteins, including DNA polymerase and replication protein. A RPA-like single-stranded DNA-binding proteins. The newly synthesized leading-plus strand remains covalently linked to the 3′-OH end of the parent-plus strand during the continuous displacement of its 5′-OH end. When the leading strand makes a complete turn, Rep catalyzes a strand-transfer reaction followed by the release of a single-stranded DNA intermediate, the parent-minus strand, and a double-stranded DNA *Helitron* composed of both the parental-plus and a newly synthesized strand (60).

Another interesting feature of *Helitron* is its ability to intercept host genes. For example, plant *Helitrons* encode RPA-like proteins, clearly derived from RPA encoded originally by the host genome (60). Given the conservation of RPA in *Helitrons*, this protein is almost certainly involved in *Helitron* transposition, presumably as a single-stranded DNA-binding protein. *Helitrons* present in sea anemone, sea urchin, fish, and frog carry EN derived from CR1-like non-LTR retrotransposons (63, 99). Again, the conservation of the EN in different *Helitrons* from different species shows that it must be necessary for the life cycle of *Helitrons*. Finally, numerous nonautonomous *Helitrons* in the corn genome harbor exon-/intron-coding portions from many different host genes (75, 90). Therefore, *Helitrons* may function as a powerful tool of evolution, by mediating duplication, shuffling, and recruitment of host genes.

Polintons. Like *Helitrons*, the third class of DNA transposons, *Polintons*, was discovered and characterized based on computational studies (65). *Polintons* are 15–20 kb long, with 6-bp TSDs and 100–1000 bp TIRs at both ends. They are the most complex eukaryotic transposons known to date. *Polintons* code for up to 10 proteins, including a family B DNA polymerase (POLB), a retroviral-like INT, an A transposase, and an adenoviral-like cysteine protease. The first three are universal for all autonomous *Polintons* identified in protists, fungi, and animals (65).

Polinton POLB belongs to a group of protein-primed DNA polymerases encoded by genomes of bacteriophages, adenoviruses, and linear plasmids from fungi and plants. POLB and its functional motifs are well defined (10, 24, 115), and their conservation in all extremely diverged *Polinton* POLBs indicates that the DNA-DNA polymerase and proofreading activities are necessary for *Polinton* transposition. The termini of *Polintons* are composed of short 1–3-bp tandem repeats, which are necessary for the slide-back mechanism in protein-primed DNA synthesis studied in bacteriophages (88). Based on these observations, it was proposed that *Polintons* propagate through protein-primed self-synthesis by POLB (65). First, during host genome replication, the INT-catalyzed excision of *Polinton* from the host DNA leads to an extrachromosomal single-stranded *Polinton* that forms a racket-like structure. Second, the *Polinton* POLB replicates the extrachromosomal *Polinton*. Finally, after the double-stranded *Polinton* is synthesized, the INT molecules bind to its termini and catalyze its integration into the host genome.

FROM TRANSPOSABLE ELEMENTS TO GENES

The first clear example of a functional protein-coding gene that evolved from a former TE was centromere protein B (CENP-B) (110, 116). CENP-B is conserved in mammals and binds a specific 17-bp site in the human α-satellite, but its exact function is still not clear. Based on gene knockout studies, it appears that CENP-B is involved in reproduction rather than in centromere-related activities, as was originally predicted (32, 112). Three yeast CENP-B homologs were reported (47), but given their low (<30%) identity to mammalian CENP-B and a similar range of identities between different transposases in mammals, plants, and fungi, they probably evolved in the yeast genome from the *Mariner/Pogo* transposase independently of the mammalian CENP-B.

Approximately 50–100 protein-coding genes in the mammalian genome evolved from coding sequences of DNA transposons and retrotransposons (12, 18, 62, 64, 66, 76, 120; V.V. Kapitonov & J. Jurka, unpublished work). Most of these genes ascended from transposases (*Mariner/Pogo*, *hAT*, *piggyBac*, *P*, *Harbinger*, and *Transib*). The RAG1 protein, which is a key player in V(D)J recombination (103, 113), is probably the most ancient known host protein derived from a transposable element (62, 64). RAG1 evolved some 500 million years ago (mya) in a common ancestor of jawed vertebrates from a *Transib* DNA transposase (64). It is also the only transposase-derived host gene with demonstrated nuclease-/transposase-like activities. Biological properties of the remaining transposase-derived genes are either not known or linked to DNA/RNA binding (52, 78, 114). The RAG1-based immune system is also the only example of a complex host machinery that evolved from transposase and TIRs from the same family of transposons (64). There are other genes that may be involved in DNA rearrangements, as they encode transposase-derived proteins sharing conserved catalytic amino acids with corresponding transposases. An example of such a conserved gene is *HARBI1*, which evolved from the *Harbinger* transposase in a common ancestor of fish, birds, frogs, and mammals (62).

Other potential sources of novel protein-coding genes are LTR retrotransposons. For instance, >50 protein-encoding genes syntenic between the human and mouse genomes evolved from the *gag* protein encoded by *Gypsy* LTR retrotransposons, which were active in ancestral genomes (12, 18, 66, 79, 94, 119). One of the *Gypsy*-derived genes, called PEG10 or KIAA1051, includes *Gypsy gag* and protease domains, which are fused together through the –1 ribosomal frame-shift mechanism typical for *Gypsy* elements (82, 94, 119). Although its exact function is still unknown, PEG10 is important for mouse parthenogenetic development based on observed embryonic lethality due to placental defects in PEG10 knockout mice (95). Although there are >30 examples of host genes evolved from DNA transposases, there is only one example of a recruited RT: the mammalian Rtl1 or PEG11 gene, which evolved from the *Gypsy gag* and RT (104). Interestingly, both PEG10 and PEG11 are paternally expressed genes, and more than 50% of all *gag*-derived genes reside on the X chromosome.

Finally, many microRNA genes appear to have evolved from TEs, and their involvement in gene regulation appears to be an outcome of the antagonistic relationship between TEs and the host genome. Expression of TEs and generation of repetitive DNA, including tandem repeats, are countered by RNA degradation and DNA methylation (17, 22, 100, 118) mediated by small RNAs (sRNAs) (~20–26 bp) generated from the targeted repetitive DNA. Analogous processes are involved in modulating chromatin structure and regulating gene expression (7, 8, 15, 22, 101). Many of such processes are mediated by sRNAs derived from evolutionarily conserved precursors (21).

Most of the epigenetic regulation of endogenous genes in *A. thaliana* appears to have evolved from mechanisms to silence TEs (128). Furthermore, some mammalian precursors of microRNAs (miRNAs) appear to be derived from ancient MIR (SINE) and *L2* (LINE) elements (108), or even younger *Alu* (SINE) elements and processed pseudogenes (25, 109). Recent evidence that 5′ *Alus* can function as RNA polymerase promoters for miRNAs (11) further supports the contributions of TEs to the origin and expression of miRNAs involved in mammalian gene regulation.

OTHER HIGHLY CONSERVED TRANSPOSABLE ELEMENTS

Recent systematic comparisons of complete genomic sequences revealed the existence of noncoding sequences that are highly conserved across multiple species (6). They include LF-SINE (5), MER121 (58), AmnSINE1, and SINE3-1 (92, 127), which are SINE, or SINE-like, elements preserved in highly diverse vertebrates from *Latimeria* and reptiles to mammals. An additional 83 families of low and moderately repeated elements were reported recently and deposited in Repbase (38, 53). The list includes 20 Eulor families, 15 newly analyzed MER families, 31 UCON families, 14 LINE-like families (X*_LINEs), and 3 MARE families. Eulor families are relatively small, with self-complementary regions suggesting that they might have been derived from DNA transposons. Likewise, many MER elements also resemble nonautonomous DNA transposons. Furthermore, mammalian-specific MARE3 is a tRNA-derived SINE (38). X*_LINEs, where the asterisk stands for specification of one of the 14 families, were directly or indirectly derived from autonomous non-LTR retrotransposons, a fact supported by significant similarities between their translatable regions to diverse LINE elements (38).

Table 2 shows densities of the above-described families of repeats, including a moderately repetitive *L4* family (36), in five vertebrate genomes. Columns 2 and 3 show densities of the same families in human conserved sequences (106) and *cis*-regulatory modules (CRMs) (9). For some families, the densities of TEs in CRMs can be as much as a factor of magnitude higher than the average human genomic density. Similar

Table 2 Densities of selected repetitive families per 1 Mb of DNA sequence

	H.s.	Cons.	CRMs	E.t.	M.d.	G.g.	X.t.
AmnSINE1_GG	0.28	0.81	0.94	0.17	0.15	0.64	0.24
AmnSINE1_HS	0.17	1.72	1.65	0.13	0.27	1.10	0.02
Eulor*	0.42	5.00	6.16	0.35	0.44	1.66	0.15
L4	1.69	3.21	1.93	0.53	0.13	0.00	0.00
LF-SINE	0.18	1.96	2.19	0.13	0.29	1.48	0.07
MARE1-2	0.96	1.45	1.23	0.35	1.21	0.00	0.00
MARE3	0.19	0.91	0.93	0.07	0.45	0.00	0.00
MER121	0.30	3.39	0.33	0.20	0.30	0.00	0.00
MER122-136	1.22	8.69	9.52	0.63	1.53	1.85	0.07
SINE3-1*	0.03	0.07	0.08	0.04	0.11	0.02	0.17
UCONS1-31	0.60	6.63	6.96	0.51	0.65	2.55	0.15
X*_LINE	0.63	2.98	3.18	0.28	1.19	1.06	0.57

DNA origin: human (H.s., *Homo sapiens*), tenrec (E.t., *Echinops telfairi*), Brazilian gray short-tailed opposum (M.d., *Monodelphis domestica*), chicken (G.g., *Gallus gallus*), and the pipid frog (X.t., *Xenopus tropicalis*). Columns 2 and 3 represent densities in conserved human sequences (106) and *cis*-regulatory modules (CRMs) (9). Asterisks indicate additional subclassification.

overrepresentation can be found in conserved regions, which partially overlap with CRMs. Many CRMs are tissue specific (9), which may be significant for understanding developmental evolution, especially in light of recent evidence that TEs can affect developmental processes in mammalian oocytes and preimplantation embryos (96). The most overrepresented repeats are broadly conserved in mammals, chicken, and other vertebrates (e.g., LF-SINE). The overrepresentation may be due to the fact that nonconserved copies continue to decay over time, leading to lower overall genomic densities. However, the proportions of mammalian-specific repeat families reveal a more complex pattern, which can be seen in five mammalian families: MARE1-2, MARE3, *L4*, and MER121. These families are moderately repetitive as their copy numbers in the human genome range from ~600 (MARE3) to ~5000 (*L4*). The density of MARE1-2 repeats is about 50% higher in conserved regions than the overall human genomic density of these elements, whereas the density of MARE3 repeats is almost five times higher both in conserved regions and CRMs. In the conserved regions, the densities of *L4* and MER121 repeats are around 2 and 10 times higher, respectively, but in CRMs they are comparable with human genomic densities. In other words, MARE3 accumulated many times faster in both CRMs and conserved regions than MARE1-2. On the other hand, the accumulation of *L4* and MER121 in CRMs was marginal at best, whereas their overrepresentation in conserved regions was substantial, particularly MER121. A potential scenario of fixation of certain families in regulatory regions is further discussed in the last section in the context of speciation.

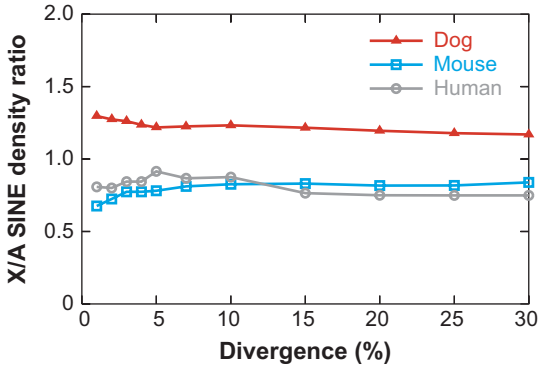

Figure 2

Ratios of short interspersed element (SINE) densities on chromosome X relative to autosomes from dog, mouse, and human.

CUMULATIVE IMPACT ON GENOMIC STRUCTURE

A systematic buildup of repetitive DNA in the genome is due to an excess of insertions over deletions of TEs. Over time it can lead to large-scale genomic changes, which were recently studied in some detail in mammalian genomes. The most prominent TEs in Euterian mammals are LINE1, (*L1*) non-LTR retroelements, and the associated nonautonoumous SINE elements. SINEs are particularly convenient for comparative studies of insertion and elimination of TEs in mammals due to their abundance, moderate lengths, and recent, or even ongoing, retrotransposition. Previous analyses of the human and mouse genomes indicate that the insertion and elimination of young SINE elements occur in male germlines (54–56). In the absence of selection, the systematic insertion of TEs in male germlines leads to their overrepresentation on chromosome Y and underrepresentation on chromosome X relative to autosomes. Analogous insertions in female germlines result in their overrepresentation on chromosome X relative to autosomes and their total absence on chromosome Y, except in regions undergoing X-Y recombination. **Figure 2** shows that proportions of SINE densities on chromosome X and autosomes (A) in human and mouse are close to 2/3, which is consistent with passing active SINEs through the male germline (54, 55). In the dog genome, the analogous X/A proportions are close to 4/3, the value predicted for transmission through female germ cells only (55, 124).

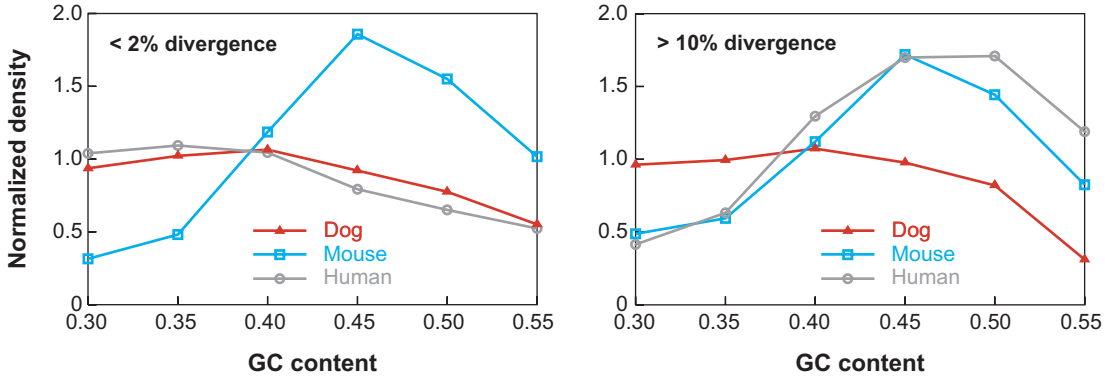

Figure 3

Normalized densities of younger (*left*) and older (*right*) short interspersed elements (SINEs) from dog (*red*), mouse (*blue*), and human (*gray*) genomes, plotted for different ranges of guanosine and cytosine (GC) content.

The biological impact of germline-specific transmission of TEs remains to be determined. Female-driven transmission could be among the initial mechanisms leading to an accumulation of *L1* elements on chromosome X, where they could be further accumulated by natural selection due to their potential involvement in spreading and amplifying the X-inactivation signal (83). Germline-specific expression is consistent with methylation patterns: Human *Alu*s are hypomethylated in sperm, whereas *L1*s are hypermethylated (45, 46, 123). Maternal and paternal methylation of repeats can provide genomic imprinting signatures guiding the epigenetic modification machineries to the imprinted regions (8, 123).

An emerging question is whether male and female modes of transposition can affect genetic stability of TEs due to the general premise that female germlines appear to be more genetically inert than male germlines (86, 89). The male-transmitted human *Alu* elements are originally prevalent in AT-rich DNA, but over time they tend to accumulate in GC-rich chromosomal regions (54, 76). In the mouse genome such an accumulation may be very fast (55). However, in the dog genome there is virtually no difference between the distribution of younger and older SINEs in DNA segments, indicating the lack of analogous postinsertion accumulation (**Figure 3**). Postinsertion changes in the distribution of SINE elements have been attributed to different SINE-SINE recombination patterns in gene-rich and gene-poor isochores (43, 54). This implies that recombinations between female-transmitted SINEs in the dog genome are much less frequent than those between male-transmitted SINEs in the human and mouse genomes. One of the outcomes of SINE-SINE recombination is the formation of hybrid SINEs that are no longer flanked by identical TSDs. The loss of TSDs over time has been observed in the human and mouse genomes, but not in the dog genome (37), which is consistent with the prediction of a low SINE-SINE recombination rate in dog. Finally, SINE-SINE recombinations can trigger the formation of long segmental duplications (51, 105). In contrast to human and mouse, very few such duplications have been found in dog (77), again in agreement with the data indicating low SINE-SINE recombination activities in dog. Nevertheless, a direct relationship between germline-specific transmission and stability of TEs remains to be shown.

POTENTIAL ROLE IN SPECIATION

In 1984, McClintock proposed that species could originate due to sudden, TE-induced chromosomal reorganizations followed by a complex genomic response (87). Recently, this idea was synthetically reviewed in the context of classical paleontological data and current knowledge of TEs (80). McClintock also noted that TEs can be activated suddenly due to abiotic stress, and concluded "that stress, and the genome reaction to it, may underlie many formations of new species" (87). Growing evidence confirms the role of stress in generating and unmasking genetic diversity (reviewed in 3). Activation of TEs under stress may produce beneficial outbursts of mutations at critical evolutionary junctions when populations face a choice between extinction and rapid change. Apart from potential contributions to speciation, activation of TEs under stress may also increase their chance of survival because host populations under stress are likely to shrink, and TEs are more likely to spread by genetic drift in small populations than in large ones. Other early ideas on the role of TEs in speciation (102) were inspired in part by the discovery of hybrid dysgenesis in fruit fly induced by the *P* element (67). The renewed interest in interspecies hybridization as a model for TE activation in speciation is stimulated by more evidence of retroelement upregulation in animal and plant species (34, 50, 74, 93). Recently, the upregulation of TEs in hybrid dysgenesis was linked to disturbance of the RNA silencing system in germ cells (57).

Any model of speciation that includes the participation of TEs must include changes in regulatory systems. The potential significance of repetitive DNA for evolution of eukaryotic regulation was first proposed by Britten & Davidson (13), before the relationship between TEs and repetitive DNA was established. Later, Wilson and colleagues (69, 125) proposed that gene rearrangements, particularly those leading to changes in regulatory regions, may account for the major organismal differences. Since then, more experimental evidence of changes in transcription regulation and their impact on organismal phenotypes at a microevolutionary level became available (8, 126). However, many regulatory changes are likely harmful or lethal, as indicated by the apparent lack of TE insertions in many developmental regulators (107). Yet, as discussed above, certain repeat families are overrepresented in CRMs and conserved regions, whereas others are not (38). This strongly indicates that fixation of conserved TEs in regulatory regions did occur in the past, although the process is not well understood. Some of these conserved repeats were derived from tRNA or 5SRNA elements, which might have contributed useful regulatory signals such as pol promoters. However, there are other tRNA-derived families, such as MIR, that are not overrepresented in conserved regions, in contrast to tRNA-derived MER133 or MARE3 (38, 53). There could be additional factors affecting the fixation, and among them coincidence with speciation is of particular interest. Such fixations need not necessarily come from large outbursts of (retro)transpositions. Instead, any TEs active at the time of speciation could have a better chance to make multiple impacts on regulatory regions, particularly those involved in developmental processes. Recent analysis of selected retropseudogenes revealed that the peaks of their retropositions appear to roughly correspond with major events in the primate phylogenetic history (26). More recently, it was demonstrated that significant genome expansion in three hybrid sunflower species is attributable to proliferation of *Gypsy*-type LTR retrotransposons (117). It remains to be seen if this type of rapid expansion can translate to multiple fixations of repeats in regulatory sites. Successfully modified regulatory sites could later undergo duplications, further increasing the proportions of conserved copies relative to the total number of genomic copies. Major changes during speciation are unlikely to be optimal or stable and may

require extensive fine-tuning over time. TEs may play another important role in such processes by affecting DNA methylation patterns and microRNA-mediated regulation.

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant 5 P41 LM006252-09. We thank Jolanta Walichiewicz and Andrew Gentles for help with editing the manuscript.

LITERATURE CITED

1. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–402
2. Arkhipova IR, Pyatkov KI, Meselson M, Evgen'ev MB. 2003. Retroelements containing introns in diverse invertebrate taxa. *Nat. Genet.* 33:123–24
3. Badyaev AV. 2005. Stress-induced variation in evolution: from behavioural plasticity to genetic assimilation. *Proc. Biol. Sci.* 272:877–86
4. Bandelt HJ, Quintana-Murci L, Salas A, Macaulay V. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. *Am. J. Hum. Genet.* 71:1150–60
5. Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, et al. 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* 441:87–90
6. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, et al. 2004. Ultraconserved elements in the human genome. *Science* 304:1321–25
7. Bernstein E, Allis CD. 2005. RNA meets chromatin. *Genes Dev.* 19:1635–55
8. Biemont C, Vieira C. 2006. Genetics: junk DNA as an evolutionary force. *Nature* 443:521–24
9. Blanchette M, Bataille AR, Chen X, Poitras C, Laganiere J, et al. 2006. Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression. *Genome Res.* 16:656–68
10. Blanco L, Salas M. 1996. Relating structure to function in phi29 DNA polymerase. *J. Biol. Chem.* 271:8509–12
11. Borchert GM, Lanier W, Davidson BL. 2006. RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol.* 13:1097–101
12. Brandt J, Veith AM, Volff JN. 2005. A family of neofunctionalized Ty3/gypsy retrotransposon genes in mammalian genomes. *Cytogenet. Genome Res.* 110:307–17
13. Britten RJ, Davidson EH. 1969. Gene regulation for higher cells: a theory. *Science* 165:349–57
14. Britten RJ, Kohne DE. 1968. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* 161:529–40
15. Brodersen P, Voinnet O. 2006. The diversity of RNA silencing pathways in plants. *Trends Genet.* 22:268–80
16. Brosius J. 1991. Retroposons–seeds of evolution. *Science* 251:753

17. Buchon N, Vaury C. 2006. RNAi: a defensive RNA-silencing against viruses and transposable elements. *Heredity* 96:195–202
18. Campillos M, Doerks T, Shah PK, Bork P. 2006. Computational characterization of multiple Gag-like human proteins. *Trends Genet*. 22:585–89
19. Cappello J, Handelsman K, Lodish HF. 1985. Sequence of Dictyostelium DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. *Cell* 43:105–15
20. Capy P, Langin T, Higuet D, Maurer P, Bazin C. 1997. Do the integrases of LTR-retrotransposons and class II element transposases have a common ancestor? *Genetica* 100:63–72
21. Carrington JC, Ambros V. 2003. Role of microRNAs in plant and animal development. *Science* 301:336–38
22. Chan SW, Henderson IR, Jacobsen SE. 2005. Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat. Rev. Genet.* 6:351–60
23. Craig NL. 1995. Unity in transposition reactions. *Science* 270:253–54
24. de Vega M, Lazaro JM, Salas M, Blanco L. 1998. Mutational analysis of phi29 DNA polymerase residues acting as ssDNA ligands for 3′–5′ exonucleolysis. *J. Mol. Biol.* 279:807–22
25. Devor EJ. 2006. Primate microRNAs miR-220 and miR-492 lie within processed pseudogenes. *J. Hered.* 97:186–90
26. Devor EJ, Moffat-Wilson KA. 2003. Molecular and temporal characteristics of human retropseudogenes. *Hum. Biol.* 75:661–72
27. Doak TG, Doerder FP, Jahn CL, Herrick G. 1994. A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. *Proc. Natl. Acad. Sci. USA* 91:942–46
28. Duncan L, Bouckaert K, Yeh F, Kirk DL. 2002. Kangaroo, a mobile element from Volvox carteri, is a member of a newly recognized third class of retrotransposons. *Genetics* 162:1617–30
29. Eickbush TH, Malik HS. 2002. Origins and evolution of retrotransposons. In *Mobile DNA II*, ed. NL Craig, R Craigie, M Gellert, AM Lambowitz, pp. 1111–144. Washington, DC: ASM Press
30. Evgen'ev MB, Arkhipova IR. 2005. Penelope-like elements—a new class of retroelements: distribution, function and possible evolutionary significance. *Cytogenet. Genome Res.* 110:510–21
31. Feschotte C. 2004. Merlin, a new superfamily of DNA transposons identified in diverse animal genomes and related to bacterial IS1016 insertion sequences. *Mol. Biol. Evol.* 21:1769–80
32. Fowler KJ, Hudson DF, Salamonsen LA, Edmondson SR, Earle E, et al. 2000. Uterine dysfunction and genetic modifiers in centromere protein B-deficient mice. *Genome Res.* 10:30–41
33. Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, et al. 2005. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and A. oryzae. *Nature* 438:1105–15
34. Garcia Guerreiro MP, Biemont C. 1995. Changes in the chromosomal insertion pattern of the copia element during the process of making chromosomes homozygous in *Drosophila melanogaster*. *Mol. Gen. Genet.* 246:206–11
35. Gardner MJ, Hall N, Fung E, White O, Berriman M, et al. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419:498–511
36. Genetic Information Research Institute. 2007. *Repbase update*. **http://www.girinst.org/repbase/index.html**

37. Gentles AJ, Kohany O, Jurka J. 2005. Evolutionary diversity and potential recombinogenic role of integration targets of non-LTR retrotransposons. *Mol. Biol. Evol.* 22:1983–91
38. Gentles AJ, Wakefield MJ, Kohany O, Gu W, Batzer MA, et al. 2007. Evolutionary dynamics of transposable elements in the short-tailed opossum Monodelphis domestica. *Genome Res.* In press. doi:10.1101/gr.6070707
39. Goodwin TJ, Butler MI, Poulter RT. 2003. Cryptons: a group of tyrosine-recombinase-encoding DNA transposons from pathogenic fungi. *Microbiology* 149:3099–109
40. Goodwin TJ, Poulter RT. 2001. The DIRS1 group of retrotransposons. *Mol. Biol. Evol.* 18:2067–82
41. Goodwin TJ, Poulter RT. 2004. A new group of tyrosine recombinase-encoding retrotransposons. *Mol. Biol. Evol.* 21:746–59
42. Grindley, ND, Sherratt DJ. 1979. Sequence analysis at IS1 insertion sites: models for transposition. *Cold Spring Harb Symp Quant Biol.* 43:1257–61
43. Hackenberg M, Bernaola-Galvan P, Carpena P, Oliver JL. 2005. The biased distribution of Alus in human isochores might be driven by recombination. *J. Mol. Evol.* 60:365–77
44. Hartl DL, Dykhuizen DE, Miller RD, Green L, de Framond J. 1983. Transposable element IS50 improves growth rate of E. coli cells without transposition. *Cell* 35:503–10
45. Hellman-Blumberg U, Hintz MF, Gatewood JM, Schmid CW. 1993. Developmental differences in methylation of human Alu repeats. *Mol. Cell. Biol.* 8:4523–30
46. Howlett SK, Reik W. 1991. Methylation levels of maternal and paternal genomes during preimplantation development. *Development* 113:119–27
47. Irelan JT, Gutkin GI, Clarke L. 2001. Functional redundancies, distinct localizations and interactions among three fission yeast homologs of centromere protein-B. *Genetics* 157:1191–203
48. Jacobs ME, Sanchez-Blanco A, Katz LA, Klobutcher LA. 2003. Tec3, a new developmentally eliminated DNA element in Euplotes crassus. *Eukaryot. Cell* 2:103–14
49. Jordan IK, Matyunina LV, McDonald JF. 1999. Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. *Proc. Natl. Acad. Sci. USA* 96:12621–25
50. Josefsson C, Dilkes B, Comai L. 2006. Parent-dependent loss of gene silencing during interspecies hybridization. *Curr. Biol.* 16:1322–28
51. Jurka J. 2004. Evolutionary impact of human Alu repetitive elements. *Curr. Opin. Genet. Dev.* 14:603–8
52. Jurka J, Kapitonov VV. 1999. Sectorial mutagenesis by transposable elements. *Genetica* 107:239–48
53. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* 110:462–67
54. Jurka J, Kohany O, Pavlicek A, Kapitonov VV, Jurka MV. 2004. Duplication, coclustering, and selection of human Alu retrotransposons. *Proc. Natl. Acad. Sci. USA* 101:1268–72
55. Jurka J, Kohany O, Pavlicek A, Kapitonov VV, Jurka MV. 2005. Clustering, duplication and chromosomal distribution of mouse SINE retrotransposons. *Cytogenet. Genome Res.* 110:117–23
56. Jurka J, Krnjajic M, Kapitonov VV, Stenger JE, Kohkanyy O. 2002. Active Alu elements are passed primarily through paternal germlines. *Theor. Popul. Biol.* 61:519–30
57. Kalmykova AI, Klenov MS, Gvozdev VA. 2005. Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. *Nucleic Acids Res.* 33:2052–59
58. Kamal M, Xie X, Lander ES. 2006. A large family of ancient repeat elements in the human genome is under strong selection. *Proc. Natl. Acad. Sci. USA* 103:2740–45

59. Kapitonov VV, Jurka J. 1999. Molecular paleontology of transposable elements from *Arabidopsis thaliana*. *Genetica* 107:27–37
60. Kapitonov VV, Jurka J. 2001. Rolling-circle transposons in eukaryotes. *Proc. Natl. Acad. Sci. USA* 98:8714–19
61. Kapitonov VV, Jurka J. 2003. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proc. Natl. Acad. Sci. USA* 100:6569–74
62. Kapitonov VV, Jurka J. 2004. Harbinger transposons and an ancient *HARBI1* gene derived from a transposase. *DNA Cell Biol.* 23:311–24
63. Kapitonov VV, Jurka J. 2005. Helitron-1_SP, a family of autonomous Helitrons in the sea urchin genome. *Repbase Rep.* 5:393
64. Kapitonov VV, Jurka J. 2005. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLOS Biol.* 3:e181
65. Kapitonov VV, Jurka J. 2006. Self-synthesizing DNA transposons in eukaryotes. *Proc. Natl. Acad. Sci. USA* 103:4540–45
66. Kapitonov VV, Pavlicek A, Jurka J. 2004. Anthology of human repetitive DNA. In *Encyclopedia of Molecular Cell Biology and Molecular Medicine*, ed. RA Meyers, pp. 251–305. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA
67. Kidwell MG, Kidwell JF, Sved JA. 1977. Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics* 86:813–33
68. Kidwell MG, Lisch DR. 2001. Perspective: transposable elements, parasitic DNA, and genome evolution. *Evol. Int. J. Org. Evol.* 55:1–24
69. King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188:107–16
70. Kojima KK, Fujiwara H. 2004. Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets. *Mol. Biol. Evol.* 21:207–17
71. Kordis D, Gubensek F. 1997. Bov-B long interspersed repeated DNA (LINE) sequences are present in Vipera ammodytes phospholipase A2 genes and in genomes of Viperidae snakes. *Eur. J. Biochem.* 246:772–79
72. Kulpa DA, Moran JV. 2006. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat. Struct. Mol. Biol.* 13:655–60
73. Kunze R, Weil CF. 2002. The hAT and CACTA superfamilies of plant transposons. In *Mobile DNA II*, ed. NL Craig, R Craigie, M Gellert, AM Lambowitz, pp. 565–610. Washington, DC: ASM Press
74. Labrador M, Farre M, Utzet F, Fontdevila A. 1999. Interspecific hybridization increases transposition rates of Osvaldo. *Mol. Biol. Evol.* 16:931–37
75. Lai J, Li Y, Messing J, Dooner HK. 2005. Gene movement by Helitron transposons contributes to the haplotype variability of maize. *Proc. Natl. Acad. Sci. USA* 102:9068–73
76. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921
77. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438:803–19
78. Liu W, Seto J, Sibille E, Toth M. 2003. The RNA binding domain of Jerky consists of tandemly arranged helix-turn-helix/homeodomain-like motifs and binds specific sets of mRNAs. *Mol. Cell Biol.* 23:4083–93
79. Llorens C, Marin I. 2001. A mammalian gene evolved from the integrase domain of an LTR retrotransposon. *Mol. Biol. Evol.* 18:1597–600

80. Lonnig WE, Saedler H. 2002. Chromosome rearrangements and transposable elements. *Annu. Rev. Genet.* 36:389–410
81. Lorenzi HA, Robledo G, Levin MJ. 2006. The VIPER elements of trypanosomes constitute a novel group of tyrosine recombinase-enconding retrotransposons. *Mol. Biochem. Parasitol.* 145:184–94
82. Lux A, Beil C, Majety M, Barron S, Gallione CJ, et al. 2005. Human retroviral gag- and gag-pol-like proteins interact with the transforming growth factor-beta receptor activin receptor-like kinase 1. *J. Biol. Chem.* 280:8482–93
83. Lyon MF. 1998. X-chromosome inactivation: a repeat hypothesis. *Cytogenet. Cell Genet.* 80:133–37
84. Lyozin GT, Makarova KS, Velikodvorskaja VV, Zelentsova HS, Khechumian RR, et al. 2001. The structure and evolution of Penelope in the virilis species group of *Drosophila*: an ancient lineage of retroelements. *J. Mol. Evol.* 52:445–56
85. Mah C, Sarkar R, Zolotukhin I, Schleissing M, Xiao X, et al. 2003. Dual vectors expressing murine factor VIII result in sustained correction of hemophilia A mice. *Hum. Gene Ther.* 14:143–52
86. Makova KD, Yang S, Chiaromonte F. 2004. Insertions and deletions are male biased too: a whole-genome analysis in rodents. *Genome Res.* 14:567–73
87. McClintock B. 1984. The significance of responses of the genome to challenge. *Science* 226:792–801
88. Mendez J, Blanco L, Esteban JA, Bernad A, Salas M. 1992. Initiation of phi 29 DNA replication occurs at the second 3′ nucleotide of the linear template: a sliding-back mechanism for protein-primed DNA replication. *Proc. Natl. Acad. Sci. USA* 89:9579–83
89. Miller W, Makova KD, Nekrutenko A, Hardison RC. 2004. Comparative genomics. *Annu. Rev. Genom. Hum. Genet.* 5:15–56
90. 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
91. Morrish TA, Garcia-Perez JL, Stamato TD, Taccioli GE, Sekiguchi J, Moran JV. 2007. Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature* 446:208–12
92. Nishihara H, Smit AF, Okada N. 2006. Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res.* 16:864–74
93. O'Neill RJ, O'Neill MJ, Graves JA. 1998. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* 393:68–72
94. Ono R, Kobayashi S, Wagatsuma H, Aisaka K, Kohda T, et al. 2001. A retrotransposon-derived gene, *PEG10*, is a novel imprinted gene located on human chromosome 7q21. *Genomics* 73:232–37
95. Ono R, Nakamura K, Inoue K, Naruse M, Usami T, et al. 2006. Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. *Nat. Genet.* 38:101–6
96. Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, et al. 2004. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* 7:597–606
97. Plasterk RHA, Van Luenen HGAM. 2002. The Tc1/mariner family of transposable elements. In *Mobile DNA II*, ed. NL Craig, R Craigie, M Gellert, AM Lambowitz, pp. 519–32. Washington DC: ASM Press

98. Poulter RT, Goodwin TJ. 2005. DIRS-1 and the other tyrosine recombinase retrotransposons. *Cytogenet. Genome Res.* 110:575–88
99. Poulter RT, Goodwin TJ, Butler MI. 2003. Vertebrate helentrons and other novel Helitrons. *Gene* 313:201–12
100. Qi Y, He X, Wang XJ, Kohany O, Jurka J, Hannon GJ. 2006. Distinct catalytic and noncatalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* 443:1008–12
101. Richards EJ. 2006. Inherited epigenetic variation: revisiting soft inheritance. *Nat. Rev. Genet.* 7:395–401
102. Rose MR, Doolittle WF. 1983. Molecular biological mechanisms of speciation. *Science* 220:157–62
103. Schatz DG, Oettinger MA, Baltimore D. 1989. The V(D)J recombination activating gene, RAG-1. *Cell* 59:1035–48
104. Seitz H, Youngson N, Lin SP, Dalbert S, Paulsen M, et al. 2003. Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nat. Genet.* 34:261–62
105. Sharp AJ, Cheng Z, Eichler EE. 2006. Structural variation of the human genome. *Annu. Rev. Genom. Hum. Genet.* 7:407–42
106. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15:1034–50
107. Simons C, Pheasant M, Makunin IV, Mattick JS. 2006. Transposon-free regions in mammalian genomes. *Genome Res.* 16:164–72
108. Smalheiser NR, Torvik VI. 2005. Mammalian microRNAs derived from genomic repeats. *Trends Genet.* 21:322–26
109. Smalheiser NR, Torvik VI. 2006. Alu elements within human mRNAs are probable microRNA targets. *Trends Genet.* 22:532–36
110. Smit AF, Riggs AD. 1996. Tiggers and DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci. USA* 93:1443–48
111. Sun FJ, Fleurdepine S, Bousquet-Antonelli C, Caetano-Anolles G, Deragon JM. 2007. Common evolutionary trends for SINE RNA structures. *Trends Genet.* 23:26–33
112. Tomascik-Cheeseman L, Marchetti F, Lowe X, Shamanski FL, Nath J, et al. 2002. CENP-B is not critical for meiotic chromosome segregation in male mice. *Mutat. Res.* 513:197–203
113. Tonegawa S. 1983. Somatic generation of antibody diversity. *Nature* 302:575–81
114. Toth M, Grimsby J, Buzsaki G, Donovan GP. 1995. Epileptic seizures caused by inactivation of a novel gene, *jerky*, related to centromere binding protein-B in transgenic mice. *Nat. Genet.* 11:71–75
115. Truniger V, Blanco L, Salas M. 1999. Role of the "YxGG/A" motif of Phi29 DNA polymerase in protein-primed replication. *J. Mol. Biol.* 286:57–69
116. Tudor M, Lobocka M, Goodell M, Pettitt J, O'Hare K. 1992. The pogo transposable element family of *Drosophila melanogaster*. *Mol. Gen. Genet.* 232:126–34
117. Ungerer MC, Strakosh SC, Zhen Y. 2006. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Curr. Biol.* 16:R872–73
118. Vaughn MW, Martienssen R. 2005. It's a small RNA world, after all. *Science* 309:1525–26
119. Volff J, Korting C, Schartl M. 2001. Ty3/Gypsy retrotransposon fossils in mammalian genomes: Did they evolve into new cellular functions? *Mol. Biol. Evol.* 18:266–70
120. Volff JN. 2006. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *Bioessays* 28:913–22

121. Volff JN, Hornung U, Schartl M. 2001. Fish retroposons related to the Penelope element of *Drosophila* virilis define a new group of retrotransposable elements. *Mol. Genet. Genom.* 265:711–20
122. Walbot V, Rudenko GN. 2002. MuDR/Mu transposable elements of maize. In *Mobile DNA II*, ed. NL Craig, R Craigie, M Gellert, AM Lambowitz, pp. 533–64. Washington DC: ASM Press
123. Walter J, Hutter B, Khare T, Paulsen M. 2006. Repetitive elements in imprinted genes. *Cytogenet. Genome Res.* 113:109–15
124. Weiner AM, Deininger PL, Efstratiadis A. 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* 55:631–61
125. Wilson AC, Sarich VM, Maxson LR. 1974. The importance of gene rearrangement in evolution: evidence from studies on rates of chromosomal, protein, and anatomical evolution. *Proc. Natl. Acad. Sci. USA* 71:3028–30
126. Wray GA. 2003. Transcriptional regulation and the evolution of development. *Int. J. Dev. Biol.* 47:675–84
127. Xie X, Kamal M, Lander ES. 2006. A family of conserved noncoding elements derived from an ancient transposable element. *Proc. Natl. Acad. Sci. USA* 103:11659–64
128. Zilberman D, Henikoff S. 2005. Epigenetic inheritance in *Arabidopsis*: selective silence. *Curr. Opin. Genet. Dev.* 15:557–62

Annual Review of Genomics and Human Genetics

Volume 8, 2007

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Genomics and Human Genetics* chapters may be found at http://genom.annualreviews.org/