Garrido-Ramos MA (ed): Repetitive DNA.
Genome Dyn. Basel, Karger, 2012, vol 7, pp 92–107

SINEs as Driving Forces in Genome Evolution

J. Schmitz

Institute of Experimental Pathology, University of Münster, Münster, Germany

Abstract

SINEs are short interspersed elements derived from cellular RNAs that repetitively retropose via RNA intermediates and integrate more or less randomly back into the genome. SINEs propagate almost entirely vertically within their host cells and, once established in the germline, are passed on from generation to generation. As non-autonomous elements, their reverse transcription (from RNA to cDNA) and genomic integration depends on the activity of the enzymatic machinery of autonomous retrotransposons, such as long interspersed elements (LINEs). SINEs are widely distributed in eukaryotes, but are especially effectively propagated in mammalian species. For example, more than a million Alu-SINE copies populate the human genome (approximately 13% of genomic space), and few master copies of them are still active. In the organisms where they occur, SINEs are a challenge to genomic integrity, but in the long term also can serve as beneficial building blocks for evolution, contributing to phenotypic heterogeneity and modifying gene regulatory networks. They substantially expand the genomic space and introduce structural variation to the genome. SINEs have the potential to mutate genes, to alter gene expression, and to generate new parts of genes. A balanced distribution and controlled activity of such properties is crucial to maintaining the organism's dynamic and thriving evolution.

SINEs are short interspersed elements whose master copies repetitively retropose via RNA intermediates. The molecular origin of SINEs can be traced back to an organism's own cellular small RNAs, especially highly expressed tRNAs and parts of the 7SL RNA, and less frequently 5S rRNA, that are equipped with their own internal RNA polymerase III (Pol III) promoter. Actively transcribed RNAs are coincidently recognized at their 3′ ends by contemporary autonomous retrotransposon-derived enzymes (e.g. from long interspersed elements, LINEs), reverse transcribed, and more or less randomly inserted into the genome [1]. The new genomic location is crucial to determining whether the new insertion becomes an active SINE (rarely) or an inactive retropseudosequence like millions of other elements. Effective subsequent transcription requires additional sequence motifs

upstream of the insertion site [2]. The transcription of, for example, new tRNA- or 7SL-derived SINEs starts 10–12 nucleotides (nt) upstream of their internal Pol III promoter box A, extends beyond their characteristic oligo(A)-tail, and occasionally continues further downstream to a random (T) stretch (TTTT or more complex terminator motifs) [3]. Continuous and frequent transcription is a first requirement towards a prospective functional master SINE element. As non-autonomous elements, their reverse transcription (from RNA to cDNA) and genomic integration depends on the activity of the enzymatic machinery of autonomous retrotransposons, such as LINEs or the LINE-derived retroposon-like transposable elements (RTEs). Thus, the second requirement is the proficient exploitation of the retrotranspositional LINE system. To 'mimic' a suitable LINE RNA target (the genuine template for LINE retrotransposition), the 3′ ends of SINEs are generally similar or even derived from the 3′ ends of the LINE mRNAs. But organisms also have ways to hinder the proliferation of such elements, so the final path to success lies in escaping the organism's epigenetic or other defense systems. Because active SINEs do not directly contribute to the organism's fitness, they usually do not fall under natural selection that might prevent the decay of the essential transcriptional recognition sequences. This and possible protective effects of neighboring gene regions influence the limited lifetime of such elements. Furthermore, SINEs live and die with their associated autonomous retrotransposons. For example, more than 140 million years ago (Mya) the mammalian LINE2/3 activity terminated; consequently, the associated mammalian-wide interspersed repeat (MIR) SINEs suffered the same fate [4].

SINEs were first reported in rodents [5] and primates [6], but their presence is well documented in many eukaryotic lineages, including mammals, reptiles, birds, fishes, insects, molluscs, and plants (reviewed in [7]). In monotremes, which represent the first mammalian divergence, the LINE2-dependent Mon1 SINEs are the predominant mobile elements. Moving along the evolutionary tree, on the lineage leading to therians (marsupials and placentals) an especially effective autonomous retrotransposon association evolved, the LINE1-SINE system, which is particularly active in placentals. This more recently evolved LINE1 retrotransposon machinery is rather unspecific in its recognition and retrotransposition of *any* available oligo(A)-tailed RNA, such as (1) mRNAs, (2) tRNA derivates, (3) the primate-specific *Alu*-SINEs derived from the 7SL signal recognition particle, or (4) more complex composite retrotransposons, such as the SVA elements in apes (composed of a $(CCCTCT)_n$ multimer – an *Alu*-like part – a variable number of tandem repeats (VNTR) – a SINE-R derived from an LTR element of the human endogenous retrovirus (HERV) – and an oligo(A) tail) [8, 9]. Thus, in the LINE1-SINE system, the retrotranspositional frequency corresponds to the quantity of available transcribed oligo(A)-tailed templates.

Although similar SINEs might emerge de novo independently in different lineages, specific SINEs or SINE families are usually restricted to respective orders,

rarely crossing order boundaries. SINEs often are divided into different subfamilies with successive, partially overlapping waves of activity [10]. Usually, however, there is just 1 LINE/SINE system active at a given time, as is the case for the LINE1-*Alu*-SINE association in higher primates. At present, ~100 or much less *Alu*-SINE loci are possibly retrotranspositionally active in primates [11, 12], which roughly corresponds to the number of retrotranspositionally competent LINEs [13]. In human, it is estimated that active, retrotransposed SINE and LINE sequences lead to new germline insertions in at least every 30th and 50th birth, respectively [13]. However, in marsupials there is strong indication that 3 different LINE/SINE systems were active at the same time [14].

Interestingly, it has been shown for humans that LINE1s (containing AT-rich sequences) and *Alu*-SINEs (GC-rich sequences) are correspondingly distributed in AT-rich heterochromatic or GC-rich euchromatic genomic regions [15]. This is surprising because the insertion mechanisms for both are identical and possess the same slight preference for 5′-TT/AAAA-3′ genomic motifs [16]. It is speculated that the resultant lack of LINEs in gene-rich regions may be due to negative selection against such large transposable elements (TEs) that carry many internal transcription factor binding sites [17]; therefore, they might interfere with the transcription of adjacent genes. Negative selection against TE insertion in euchromatic regions may also counteract the deleterious effects triggered by TE-induced ectopic recombination [18]. Recent investigations show that, for example, the insertion preference of *Alu*-SINEs differs between old (insertion in GC-rich regions) and young (insertion in AT-rich regions) elements [19].

The effects of SINEs on genomes are wide-ranging, and the present review summarizes some of the most important aspects of SINEs as substrates for evolution.

SINEs Substantially Expand the Genomic Space

Because retroposition involves the copying of sequences and insertion of the copy back into the genome, SINEs and other TEs have the potential to substantially increase the size of genomes. One of the most impressive examples of genomic expansion is the doubling of the maize genome in just ~5 million years via proliferation of TEs [20]. Also in mammals the contribution of TEs to genome size is substantial; up to half of the mammalian genome is derived from recognizable TEs. Fortunately, only a small number of such insertions remain active. Most SINE copies and other retroposed RNAs as well, remain as inactive parts of the genome and are never transcribed directly. For example, only a few of the million *Alu*-SINE genomic copies are still actively transcribed [11] and contribute further to the continuous increase in genomic size. Below are 2 of the more interesting examples of SINEs or SINE-like elements recently discovered that have contributed to increases in genome sizes.

Example of Genome Extension by Tailless Retropseudogenes

As indicated above, the LINE system is not only responsible for its own proliferation, but also that of the SINEs and other small nuclear RNAs. In particular, over the last 140 million years, the special LINE1 retrotransposon has produced an enormous number of fragmented RNA genomic insertions known as tailless retropseudogenes [21]. Such tailless RNA copies are specifically truncated at structural loop regions (e.g. in tRNA loops or in other single stranded regions), and the insertion of the cDNAs is directed by sequence complementarity of their 3′-terminal 2–18 nucleotides and corresponding genomic loci. However, this variant insertion mechanism does not necessarily exclude the initial recognition of an oligo(A)-tail by LINE1 and retroposition via internal priming. Because tailless retropseudogenes were so far only found in therian mammals, their distribution by the LINE1 retrotranspositional system is obvious.

Example of Genome Extension by RTE-snoRNA Retroposition

The retrotranspositional process can be very 'inventive' in expanding the genome size, as demonstrated by the procedure that is called RTE-snoRNA retrotransposition [22]. In platypus, an intronic small nucleolar RNA (snoRNA) housekeeping gene (that is normally co-transcribed and processed with its host protein-coding gene to modify ribosomal RNAs) has been fused with the 3′-tail of a BovB_Plat RTE (bovine B platypus retrotransposon-like non-LTR transposable element). Although snoRNA distributions are usually limited to a single or few copies due to their distribution via rare duplication events and the adjacent RTE fragment can no longer be independently transcribed, together they form a formidable, genome-expanding element. The snoRNA endows the RTE fragment with the Pol II transcription of the host gene and the subsequent snoRNA-specific processing, and the RTE fragment provides the processed snoRNA with the RTE-tail necessary for co-opting the retrotranspositional machinery encoded by the active BovB_Plat autonomous element. Combined these 2 elements constitute an extremely efficient cooperation that produced more than 40,000 genomic copies in platypus [23]. Several of them are still actively expressed and at least 1 of them performs the housekeeping functions of the snoRNA, suggesting the possible functionality of many of the others also. This chimera is also present in echidna that diverged from the platypus lineage 17 Mya (unpublished data), providing strong evidence for its evolutionary conservation.

In mammals, SINEs or SINE-like elements expand the genome size significantly. They occupy ~13% of the human, ~8% of the mouse, ~10% of the opossum, and ~22% of the platypus genome (reviewed in [24]). While to date no chicken-specific SINEs have been detected, the zebra finch genome contains a few thousand CR1 LINE-mobilized SINEs (~0.03% of the genome). In reptiles, BovB LINE-mobilized SINEs are widespread and more than 100,000 copies are estimated to be present in the lizard genome (~2–5% of the genome) [25, 26]. SINEs were also detected in many fishes, where the copy number can vary substantially among different taxa. The fugu genome contains just a few thousand copies [24], while ~10% genomic SINE coverage

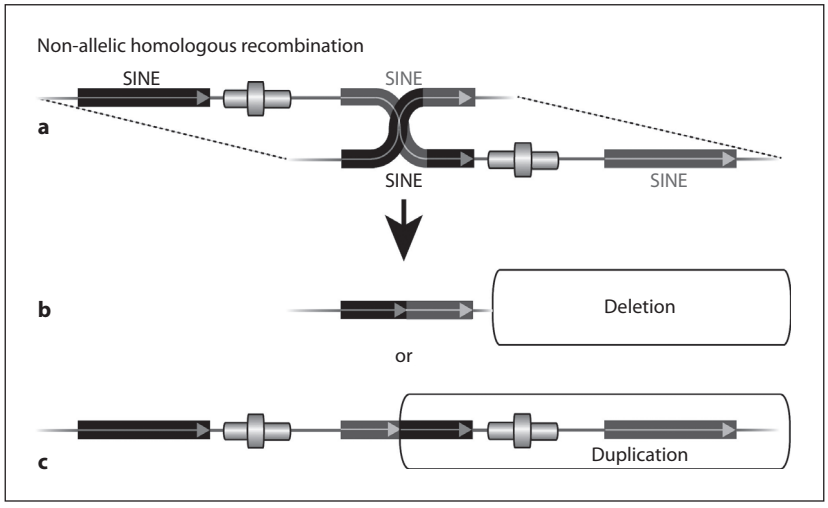

Fig. 1. SINE-induced non-allelic homologous recombination. **a** The illustration shows 2 different SINEs (black and grey bars) with high sequence similarities and embedded arrows indicating their orientations. The homologous allelic regions are shifted (dotted lines), and non-homologous (non-allelic) SINEs lead to recombination. The similar SINEs match, break, and rejoin, resulting in either a deletion of the SINE enclosed sequence region including the gene (**b**) or a larger recombined sequence carrying 2 versions of the SINE enclosed sequence which leads to a duplication of the enclosed gene (indicated by the single-exon (large grey cylinder) flanked by 2 untranslated regions (small grey cylinders) (**c**). The duplicated/deleted sequence information is framed.

is estimated for the zebrafish genome [27]. SINEs are also known from diverse invertebrate phyla, including Tunicata, Mollusca, Plathelminthes, Echinodermata, Arthropoda, and Nematoda, and plant genomes (for an overview, see [7]).

SINEs Introduce Structural Variants of the Genome

When retroposed SINEs are frequent inhabitants of the genome, they can cause sequence rearrangements. Sequence rearrangements induce variation by duplicating, deleting, or inverting sequence information. Duplicated genes can adopt the original function or acquire variant and subsequently novel functions, thus generating novel genes or multigene families. Furthermore, duplicated sequence regions constitute hot spots for subsequent segmental duplications. Because such rearrangements occur randomly, their outcome is not necessarily advantageous and they frequently lead to serious genetic disorders [28] that will eventually be selected against at the individual or population level. Any repetitive sequence block is a potential initiation point for segmental duplication. Because of their high abundance, TEs such as SINEs are frequently involved in structural variations via non-allelic homologous recombination (fig. 1).

Example TE-Associated Human Intraspecific Structural Variation

From a comparison of 2 human individuals, Xing et al. [29] determined that 706 of 8,000 intraspecies structural variants detected were induced by transposed elements, resulting in 305 kb of additional and 126 kb of removed genomic sequences. It has been shown that many segmental duplications are flanked by *Alu* elements, a strong indication for their impact on the duplication process [30]. Compared to lemurs, the human genome is expanded by ~15–20%, and 90% of the expansion is due to retroposition [31]. Some of the recent retrotransposon induced rearrangements led to numerous human diseases [28, 32]. From this analysis, Xing et al. [29] described 4 different processes of mobile element-mediated rearrangements: canonical retrotransposon insertions, non-canonical insertions associated with double-strand break repair, TE-mediated non-allelic homologous recombination leading to insertions or deletions, and non-homologous end-joining-mediated deletion via 1–7-nt 'microhomology' between TE internal breakpoints. However, structural variation is different in hetero- versus euchromatic regions. As mentioned above, TEs are often excluded (counterselected at the individual or population level) from recombination-sensitive functional (gene-rich) regions and accumulate more frequently in low-recombination-rate genomic areas, thereby maintaining genomic integrity [33]. Thus, while structural variants are selected against in euchromatic regions, because many of them cause disease, the few that are established in euchromatic regions (that do not cause disease) have a high potential to lead to novelties. The low abundance of LINEs and young *Alu*-SINEs in gene-rich regions is an example of potential selection against deleterious effects of such elements [19].

SINEs Influence or Regulate Gene Expression

Jordan et al. [34] showed that ~25% of investigated human promoter regions (~500 bp upstream of annotated transcription start sites) contain TEs, and they proposed that these elements are frequently involved in the regulatory network of transcription. SINEs and other transposed sequences control gene expression through multiple ways: by integrating into transcriptional control regions of genes, providing binding sites for transcription factors, and serving directly as enhancers or silencers of gene expression (fig. 2a). In so doing, they modulate the genomic landscape by contributing regulatory motifs. SINEs directly influence transcription and/or post-transcriptional modification in several ways. Pre-transcriptionally, they have the ability to temporally and spatially lock or unlock gene transcription during developmental stages by changing the chromatin status from a heterochromatic to a gene-activating, euchromatic structure [35]. When TEs are transcribed from the opposite strand of a gene [36], they interfere with the expression of the complementary hnRNA due to the collision of convergent RNA polymerases (fig. 2b). Post-transcriptionally, SINEs influence modification by RNA interference (RNAi). After endonucleolytic cleavage and

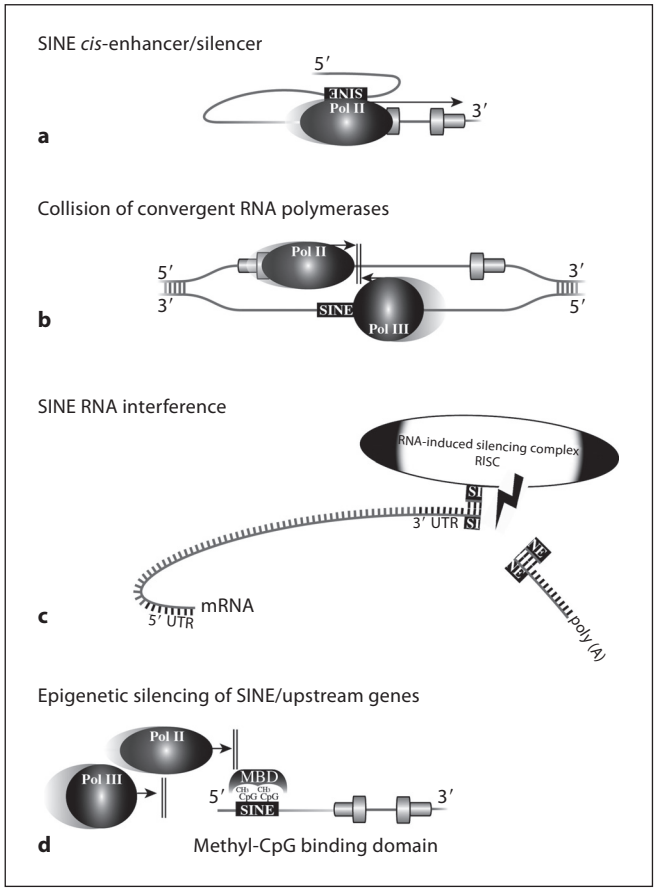

Fig. 2. SINE-induced variation of gene regulation. **a** SINE *cis*-enhancer/silencer. The upstream-located SINE (black box, written in reverse) interacts directly with the polymerase II (Pol II) promoter to enhance or silence the transcription of the associated gene. **b** SINE-induced collision of convergently moving RNA polymerases. The protein-coding gene and the SINE are located on opposite DNA strands and are transcribed in parallel by Pol II and Pol III, respectively. Due to steric inhibition of the 2 polymerases (collision), the transcription of the gene and SINE are interrupted (indicated by a double line). **c** SINE-induced RNA interference. A transcribed SINE RNA binds to a complementary sequence (e.g. a similar SINE located in the 3′ UTR of an mRNA). The double-stranded RNA region attracts the host-specific RNA-induced silencing complex (RISC) and the mRNA is fragmented (flash symbol). **d** Epigenetic silencing of a SINE and the upstream-located gene. The methyl-CpG binding domain (MBD) binds to methyl groups (CH_3) on CpG dinucleotides within the SINE sequence. The SINE-specific transcription by Pol III and the downstream-located Pol II genes are blocked (indicated by double lines). Genes in the 5′ to 3′ orientation are indicated by large cylinders (exons) and small cylinders (untranslated regions).

removal of the 3′ poly(A) tail protection, the mRNA is exposed to the cellular mRNA degradation (fig. 2c). Upon stress-induced up-regulation of SINE transcription (e.g. increased Pol III transcription after viral infection or stress-induced demethylation of CpGs), SINE RNAs can directly bind to and repress RNA Pol II transcription of mRNAs (*trans*-regulation [37]). SINEs can also serve as competitive/supportive

promoters of RNA Pol II transcription [38]. Finally, the TE-targeted methylation of CpG dinucleotides, a part of the epigenetic defensive system used to silence retrotranspositional products, can also influence the transcriptional activity of adjacent genes (fig. 2d) [39]. Since repetitive elements in the human genome contain more than 50% of these CpG dinucleotides [40], this effect can be substantial.

One should note that predominantly old elements and element families are involved in regulatory networks [41], which agrees with our own observation that predominantly old SINE insertions (see below) became persistently exonized. The following are a few examples of ultraconserved SINE modules that became exonized and acquired important cellular tasks.

Example LF-SINEs

It was recently shown that ultraconserved LF (living fossil)-SINEs, which were established ~410 Mya, eventually contribute a variety of different protein-coding cassettes and functionally conserved regulatory elements. One example is an ultraconserved enhancer derived from an LF-SINE that is located ~0.5 Mb upstream of the neurodevelopmental insulin gene enhancer gene (*ISL1*), a gene that is expressed in developing motor neurons and well conserved in tetrapodes [42]. These LF-SINEs are highly conserved non-protein-coding elements that were active in the distant past; however, some still active LF-SINEs were detected in the coelacanth, the well-known 'living fossil' that is name-giving for these elements.

Example AmnSINE1s

AmnSINE1s originated 300 Mya in the common ancestor of amniotes (reptiles, birds, and mammals) and, similar to LF-SINEs, are thought to provide building blocks involved in ancient morphological innovations [43]. They were thought to have been effectively selected by genetic drift following drastic population bottlenecks induced by the rapid change in atmospheric oxygen concentration during the Permian-Triassic era. This change probably led to the mass extinction that occurred 250 Mya [44], with few survivors carrying possibly critical genomic changes from ultraconserved SINE insertions to adapt to the altered atmospheric oxygen concentration.

More than 100 AmnSINE1 loci are conserved in mammalian genomes and are thought to convey essential mammalian-specific features mainly by influencing brain-specific transcription. That SINEs are able to influence brain-specific expression and permit functional or behavioral changes was first demonstrated with BC1, a rodent-specific tRNA-derived SINE that was exapted to brain-specific tasks [45]. Furthermore, BC1 is thought to be the master gene of thousands of related ID SINEs in rodents [46].

Example CORE-SINEs

In addition to LF-SINEs and AmnSINE1s, another ancient group of SINEs, the so-called CORE-SINEs (including MIRs, the monotreme-specific Mon1, placental Ther1,

Ther2, and the marsupial-specific, recently active MAR1s [47]) evolved important mammalian features. Approximately 10 kb downstream of the proopiomelanocortin gene (*POMC*), Santangelo et al. [48] found a specific CORE-SINE that regulates the gene activity for producing important hormone precursors in all mammalian species. The brain-specific CORE-SINE enhancer is ultraconserved in all mammals but absent in other vertebrates where regulation is thought to be performed by different regulatory elements [48].

SINEs Evolve into Protein-Coding Sequences

Sela et al. [49] found that, although retrotransposon insertion is random, ~60% of retroposed sequences in human and mouse are located in intronic regions that comprise just ~24% of the genome. Some of these well-localized retrotransposons (close to protein-coding sequences) evolve into protein-coding sequences via exonization and might subsequently acquire function in a process called exaptation [50]. The original non-protein-coding modules are usually short (up to a few hundred nucleotides) and can be included in adjacent exons via low degrees of alternative splicing without introducing shifts in open reading frames. Comparable to duplicated genes, such alternative splice products evolve mainly unconstrained because the original, dominant variant ensures functionality. If by chance they evolve an advantageous feature, natural selection improves the initially cryptic alternative splice site by selecting advantageously mutated splice sites, thereby increasing the proportion of the alternative products. In some cases, the optimization leads to the complete replacement of the original variant and constitutive expression of the new form. Evolutionary time is crucial for such a process, and we showed that from the insertion of a new SINE to its functional exonization and exaptation additional mutagenic steps are necessary, and tens or even hundreds of millions of years can elapse [51]. In principle, any suitably positioned stretch of DNA might get exonized, but reverse-oriented SINEs, such as *Alu*s [52] or MIRs [51], appear to be almost predestined for this because they are also coincidentally equipped with splice sites, comprising a 3′ AG-cryptic splice site and a preceding oligo(T) stretch – the complement of an oligo(A) tail that serves as poly-pyrimidine tract in splicing (fig. 3). There is no common codon pattern for exonized SINE cassettes. For example, the 19 exonized LF-SINEs so far analyzed used all 3 possible reading frames [42].

We think that the actual sequence of additional amino acids introduced into the protein-coding sequence by the exonized SINEs is not as significant for the evolution of the gene as is the space that they introduce, possibly optimizing protein structures by separating active protein domains to more advantageous distances from one another. An example might be the constitutively expressed MIR-cassette within the zinc finger protein 639 gene (*ZNF639*) [51].

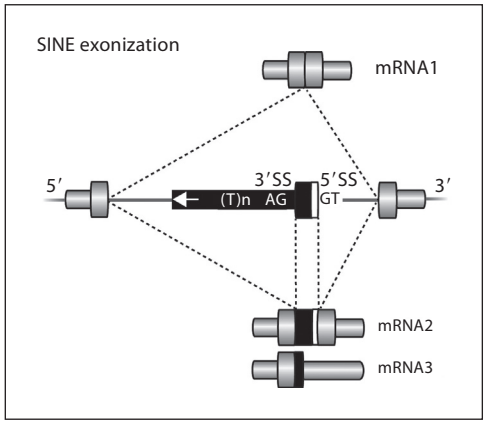

Fig. 3. SINE-induced exonization. The illustration in the middle represents the gene after an intronic reverse insertion of a SINE element. The original mRNA (without exonization) is shown at the top and the inclusion forms at the bottom. A SINE (black bar) inserts in the reverse orientation (white arrow) into an intron (thin grey line) of a gene (small cylinders = 5′ and 3′ UTRs, large cylinders = exons). The cellular splicing system is attracted by an internal 3′ cryptic AG splice site in the reverse SINE and by the poly-pyrimidine-like branch site (reverse of the SINE A-rich tail; $(T)_n$). The necessary 5′ GT splice site is randomly selected from the adjacent intron. The new, alternatively spliced exon comprises a part of the SINE (thick black box) as well as sequences of the exonized intronic region (white bar) and yields the processed mRNA2. Initially, the original mRNA1 (without the SINE exon) is predominantly processed, ensuring functionality. The new splice variant is further shaped by random changes and natural selection. If a new advantageous splice variant results, it can completely replace the original splicing product and lead to constitutive expression of the SINE cassette. However, in some instances, the new exon cassette might interrupt the open reading frame and lead to a truncated mRNA3 that is exposed to nonsense-mediated mRNA decay. The dotted lines indicate the alternatively skipped intronic regions of the mature mRNAs.

Example ZNF639

ZNF639 binds DNA and may function as a transcriptional repressor. More than 160 Mya, before the monotremes diverged from the mammalian lineage, a LINE2-mobilized MIR element (CORE-SINE) inserted into intron 5 of *ZNF639* and was fixed in the germline of the common ancestor of all mammals. The exact time of insertion is difficult to determine but occurred somewhere between the divergence of sauropsids and mammals and the first split of mammals (~300–160 Mya). However, before the mammals diversified, the element was already exonized and contributed a new exon 6 to *ZNF639*. Today, this exon is present and constitutively expressed in all mammalian species. We showed that the selection pressure on the 45 additional amino acids is moderate (K_a/K_s value 0.19) compared to the 9 zinc finger domains of the adjacent exon 7 (K_a/K_s value 0.03) [51]. This indicates that the sequence of the new amino acids is possibly not the crucial improvement, but rather, that the spacer function of the new exonized cassette, separating the protein domains of the 5th and 7th exon, is the more decisive innovation. Comparing the mammalian *ZNF639* to other vertebrate homologs provided further support

for this idea. As the classical MIRs are mammalian-specific, frogs and all other non-mammalian *ZNF639*s lack the additional MIR-coding cassette. Surprisingly, in all bird clades we also found another additional 6th exon. Moreover, the number of additional amino acids in the birds' exonized sequences is identical to that in the exonized MIR-coding cassette in mammals, but the amino acid composition is completely different, and in the case of the birds can not be assigned to any known TE. This suggests that a selectively advantageous spacer evolved independently in the 2 amniote lineages. And this is not the only example in which an exonized sequence seems to position flanking protein domains at more functionally favorable distances.

Example ADAR2

Double-strand RNA-specific editase 1 (*ADAR2/RED1*) is involved in the editing of precursor mRNAs by site-specific conversion of adenosine to inosine. Two splice variants are known for the human gene (*hADAR2*), one of which includes an *Alu*J exonization (40 amino acids) in the center of the deaminase domain (active core of the protein). The additional amino acids contribute an extra loop within 2 ß-strands of the adjacent amino acids and increase the distance between the 2nd and 3rd putative Zn^{2+}-chelating amino acids in the deaminase domain. Both variants have the same substrate specificity, but the catalytic activity of the *Alu*J inclusion variant is different, and additional potential protein interactions and regulatory functions for the enlarged splice variant have been discussed [53].

Usually it is a long obstacle course from insertion to the evolution of a novel advantageous exon function. Most exonized genes lose their additional protein-coding sequences (splice variants) by randomly accumulated mutations before an improving or novel function with a selective advantage is established. The long path to a new exon can be very versatile and may include editing processes and changes at both the DNA and the RNA levels [54].

Example NARF – Exonization via Editing

More than 43 Mya, in the common ancestor of higher primates 2 independent tail-to-tail *Alu*Sx element insertions occurred in intron 7 of the gene encoding the nuclear prelamin A recognition factor (*NARF*), which binds and processes the carboxyl-terminal tail of prelamin A. The reverse-oriented, 3′-located element already carried a C-to-GT change that later provided the 5′ splice site of a new exon 8. At the latest, sometime in the common ancestor of Great Apes, several additional changes were introduced into the precursor mRNA by adenosine-to-inosine (A-to-I) RNA editing that generated a functional 3′ splice site (AA-to-AI editing; AI functions as an AG splice site). Such editing also converted an internal UAG stop codon to a UIG codon (functions as UGG tryptophan codon). These changes were facilitated by the back-folding of the 2 adjacent *Alu*Sx elements in the unspliced pre-mRNA, building a partial double-stranded RNA structure which is a

signal for the editing enzymes to introduce inosine at adenosine positions. Finally, in the common ancestor of chimpanzee and human a TGA-to-CGA mutation at the DNA level facilitated the uninterrupted open reading frame of the new alternatively spliced exon 8 [55]. Although RNA editing is rarely involved in exonization, this example shows how evolution can act at the DNA and RNA levels to provide variations for natural selection. The process of editing is known to not only facilitate new splice variants, but is also involved in preventing additional aberrant exons [56].

In a review of the data, Sorek [57] summarized that more than 90% of new human exons were derived from repetitive sequences (~3,400 cases), with a clear dominance of *Alu*-SINE cassettes (~62%). The preexisting cryptic 3′-AG splice sites and the oligo-pyrimidine tracts of reverse SINEs, such as *Alu*s and MIRs, provide 'prefabricated' functional modules for exonization. Thus, more than 80% of exonized *Alu* elements and 60% of MIR exonizations have occurred in elements integrated in the reverse orientation [51]. In these cases, the second necessary splice sites (5′-GT) were randomly selected within the retrotransposon or taken from the adjacent intronic region, thereby including intronic sequence stretches in the new exon (fig. 3).

Down-Regulation of SINE Activity

It has been shown that SINEs, as well as other repetitive sequences, play an extraordinary role in genome evolution. Today we can trace the successes of and innovations brought about by SINEs that have already been selected, but we have only an incomplete imagination of all the past blind alleys and individual disasters caused by SINE activity. Similarly to managing a wild animal in the zoo, organisms must permanently control the activity of TEs and suppress their uncontrolled spread. This goal cannot always be met and may then lead to retrotransposon-induced diseases; whereupon, the regulatory challenge is met by purifying selection at the individual or population level. Some of the most important regulatory mechanisms to tame TE activity are described in the following section.

Example DNA Methylation (Transcriptional Gene Silencing)

In eukaryotic cells epigenetic silencing mechanisms regulate the activity of transposable elements by labeling (methylation via methyltransferase) cytosine residues at CpG dinucleotide sites and subsequently binding methyl-CpG proteins that block transcription. As mentioned previously, more than 50% of all human CpG sites are associated with transposed elements, and *Alu*-SINEs contain about 30% of the total genomic CpG sites [58].

This methylation labeling co-evolved very efficiently with the reviving activity of LINE1s in therian mammals and is also well distributed in plants. The methylation status of a genomic sequence can be copied and thus inherited upon genome replication. Because retroposition is assumed to occur predominantly in germ and

embryonic cells, it is epigenetically silenced by CpG methylation in most somatic cells (reviewed in [59]). Ongoing genomic analyses of somatic cells will soon provide a more comprehensive understanding of the retrotransposon insertion frequency in gametes vs. somatic cells.

Example RNAi (Post-Transcriptional Gene Silencing)
In general, RNAi is involved in controlling the activities of genes. The ~20-nt-long microRNAs and small interfering RNAs are the principle components for silencing mRNA and other RNAs, and especially in defending the host cell from intruding viral or other transposed elements. Expressed LINE1 sense and antisense promoters that are 'co-inserted' with the LINE in the 5′ UTR region of genes (similar to the SINE in fig. 2c), or transcribed independently from a host gene, lead to double-stranded RNA that is processed by the RNAse III homolog DICER (an endoribonuclease that cleaves double-stranded RNA) to generate short interfering RNAs. Such RNAs are induced into the RNAi-induced silencing complex (RISC), which leads to the endonucleolytic cleavage of, for example, LINE1 mRNA [60] and consequently to the silencing of LINE activities and their LINE-dependent SINE mobilization.

Example Piwi Protein and piRNA-Interacting RNA Silencing
In vertebrates, Piwi silencing activity is restricted to the germline and is mediated by an animal-specific subclass of Argonaute proteins (Piwi proteins). In zebrafish, expression was detected in both male and female gonads, but in mammals, Piwi proteins are only effective in the male germline and are associated with the ~29-nt-long piRNAs that guide the silencing machinery to bind and cleave homologous RNAs. Because many piRNAs are derived from mobile elements, it is expected that they are mainly involved in post-transcriptional gene silencing of retrotransposons and other TEs [61].

Summary

SINEs are only one component of transposed genomic elements but in many organisms, such as human, at least numerically, they predominate. A significant genomic presence of SINEs provides a pool of evolutionary building blocks that might contribute directly or indirectly to an organism's fate. The reshuffling of genomic regions induced by repetitive modules is one way of influencing affected genes by partial or complete duplication or deletion. Such significant genomic interference exposes the genome to strong purifying selection and for the organism or its next generation can mean anything from perdition to innovation. Novel properties might be suitable for adaptive survivability, especially if the environment changes drastically. Also, some indirect regulatory effects of SINEs influence gene regulatory networks by contributing expression enhancers or silencers, which evolved especially efficiently in the extraordinarily retrotranspositionally active mammalian genomes.

When located close to protein-coding genes, parts of SINEs frequently evolve into protein-coding sequences. SINE-internal sequence stretches in the reverse orientation resemble splice-like components and promote the conversion to protein-coding sequences, perhaps contributing to the optimization of special structural properties in the derived proteins. Time is obviously a crucial factor in generating and establishing evolutionary novelties. Many young exonizations or regulatory changes have not had enough time to evolve advantages to be significantly favored by natural selection and are often lost in certain populations, lineages, or species. A good example are the primate-specific *Alu*-SINE exonizations that were present in some older primate lineages but subsequently lost in younger ones. By contrast, more ancient exonizations, such as those derived from the more than 160-million-year-old MIR elements, had enough time to lead to persistent exonization or even constitutive expression of new exons in all mammalian lineages. Similarly, ultraconserved SINE sequences that have survived nearly unchanged for millions of years facilitated the evolution of essential regulatory modules. Especially we primates are much influenced by SINEs for better or worse because of the extraordinarily efficient activity and distribution of *Alu*s, which presents both a challenge for an operable genome and, at the same time, a chance to evolve novelties in the struggle of species survival.

Acknowledgements

Comparative surveys are only possible with the help of the enormous efforts of the many genome sequencing groups and colleagues that provide the tools to understand the informative content of the genome needed to unravel the secrets of mobile elements. I would like to thank Jürgen Brosius, Carsten A. Raabe and Gennady Churakov for valuable comments relating to the manuscript and Marsha Bundman for assistance in editing the information into a hopefully interesting context. Most of our own contributions to understanding the dynamics and influences of mobile elements were supported by the Deutsche Forschungsgemeinschaft (SCHM1469) and the Medical Faculty of the University of Münster.

References

1 Singer MF: SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell 1982;28:433–434.
2 Ludwig A, Rozhdestvensky TS, Kuryshev VY, Schmitz J, Brosius J: An unusual primate locus that attracted two independent Alu insertions and facilitates their transcription. J Mol Biol 2005;350:200–214.
3 Kramerov DA, Vassetzky NS: SINEs. Wiley Interdiscip Rev RNA 2011;2:772–786.
4 Smit AF, Riggs AD: MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. Nucleic Acids Res 1995;23:98–102.
5 Kramerov DA, Grigoryan AA, Ryskov AP, Georgiev GP: Long double-stranded sequences (dsRNA-B) of nuclear pre-mRNA consist of a few highly abundant classes of sequences: evidence from DNA cloning experiments. Nucleic Acids Res 1979;6:697–713.
6 Houck CM, Rinehart FP, Schmid CW: A ubiquitous family of repeated DNA sequences in the human genome. J Mol Biol 1979;132:289–306.
7 Kramerov DA, Vassetzky NS: Short retroposons in eukaryotic genomes. Int Rev Cytol 2005;247:165–221.

8 Hancks DC, Goodier JL, Mandal PK, Cheung LE, Kazazian HH Jr: Retrotransposition of marked SVA elements by human L1s in cultured cells. Hum Mol Genet 2011;20:3386–3400.
9 Wang H, Xing J, Grover D, Hedges DJ, Han K, et al: SVA elements: a hominid-specific retroposon family. J Mol Biol 2005;354:994–1007.
10 Quentin Y: The Alu family developed through successive waves of fixation closely connected with primate lineage history. J Mol Evol 1988;27:194–202.
11 Kapitonov VV, Pavlicek A, Jurka J: Anthology of human repetitive DNA; in Weinheim RAM (ed): Encyclopedia of Molecular Cell Biology and Molecular Medicine. Wiley-VCH, 2004, pp 251–305.
12 Comeaux MS, Roy-Engel AM, Hedges DJ, Deininger PL: Diverse *cis* factors controlling *Alu* retrotransposition: what causes *Alu* elements to die? Genome Res 2009;19:545–555.
13 Kazazian HH Jr: Mobile elements: drivers of genome evolution. Science 2004;303:1626–1632.
14 Nilsson MA, Churakov G, Sommer M, Tran NV, Zemann A, et al: Tracking marsupial evolution using archaic genomic retroposon insertions. PLoS Biol 2010;8:e1000436.
15 Smit AF: Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev 1999;9:657–663.
16 Jurka J: Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci USA 1997;94:1872–1877.
17 Nigumann P, Redik K, Matlik K, Speek M: Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics 2002; 79:628–634.
18 Graham T, Boissinot S: The genomic distribution of L1 elements: the role of insertion bias and natural selection. J Biomed Biotech 2006;2006:75327.
19 Belle EM, Webster MT, Eyre-Walker A: Why are young and old repetitive elements distributed differently in the human genome? J Mol Evol 2005; 60:290–296.
20 SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, et al: Nested retrotransposons in the intergenic regions of the maize genome. Science 1996;274:765–768.
21 Schmitz J, Churakov G, Zischler H, Brosius J: A novel class of mammalian-specific tailless retropseudogenes. Genome Res 2004;14:1911–1915.
22 Schmitz J, Zemann A, Churakov G, Kuhl H, Grutzner F, et al: Retroposed SNOfall–a mammalian-wide comparison of platypus snoRNAs. Genome Res 2008;18:1005–1010.
23 Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, et al: Genome analysis of the platypus reveals unique signatures of evolution. Nature 2008;453:175–183.
24 Mandal PK, Kazazian HH Jr: SnapShot: vertebrate transposons. Cell 2008;135:192–192.e1.
25 Piskurek O, Austin CC, Okada N: Sauria SINEs: novel short interspersed retroposable elements that are widespread in reptile genomes. J Mol Evol 2006;62:630–644.
26 Alfoldi J, Di Palma F, Grabherr M, Williams C, Kong L, et al: The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature 2011;477:587–591.
27 Izsvak Z, Ivics Z, Garcia-Estefania D, Fahrenkrug SC, Hackett PB: DANA elements: a family of composite, tRNA-derived short interspersed DNA elements associated with mutational activities in zebrafish *(Danio rerio)*. Proc Natl Acad Sci USA 1996;93:1077–1081.
28 Callinan PA, Batzer MA: Retrotransposable elements and human disease. Genome Dyn 2006;1: 104–115.
29 Xing J, Zhang Y, Han K, Salem AH, Sen SK, et al: Mobile elements create structural variation: analysis of a complete human genome. Genome Res 2009; 19:1516–1526.
30 Bailey JA, Liu G, Eichler EE: An *Alu* transposition model for the origin and expansion of human segmental duplications. Am J Hum Genet 2003;73:823–834.
31 Liu G, Zhao S, Bailey JA, Sahinalp SC, Alkan C, et al: Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. Genome Res 2003;13:358–368.
32 Kazazian HH Jr: Mobile elements and disease. Curr Opin Genet Dev 1998;8:343–350.
33 Bartolome C, Maside X, Charlesworth B: On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. Mol Biol Evol 2002;19:926–937.
34 Jordan IK, Rogozin IB, Glazko GV, Koonin EV: Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet 2003;19:68–72.
35 Lunyak VV, Prefontaine GG, Nunez E, Cramer T, Ju BG, et al: Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. Science 2007;317:248–251.
36 Conley AB, Miller WJ, Jordan IK: Human cis natural antisense transcripts initiated by transposable elements. Trends Genet 2008;24:53–56.

37 Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, et al: Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. Mol Cell 2008;29:499–509.
38 Ferrigno O, Virolle T, Djabari Z, Ortonne JP, White RJ, Aberdam D: Transposable B2 SINE elements can provide mobile RNA polymerase II promoters. Nat Genet 2001;28:77–81.
39 Slotkin RK, Martienssen R: Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 2007;8:272–285.
40 Xie H, Wang M, Bonaldo Mde F, Rajaram V, Stellpflug W, et al: Epigenomic analysis of Alu repeats in human ependymomas. Proc Natl Acad Sci USA 2010;107:6952–6957.
41 Gentles AJ, Wakefield MJ, Kohany O, Gu W, Batzer MA, et al: Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. Genome Res 2007;17:992–1004.
42 Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, et al: A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature 2006;441:87–90.
43 Nishihara H, Smit AF, Okada N: Functional noncoding sequences derived from SINEs in the mammalian genome. Genome Res 2006;16:864–874.
44 Okada N, Sasaki T, Shimogori T, Nishihara H: Emergence of mammals by emergency: exaptation. Genes Cells 2010;15:801–812.
45 DeChiara TM, Brosius J: Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content. Proc Natl Acad Sci USA 1987;84:2624–2628.
46 Kim J, Martignetti JA, Shen MR, Brosius J, Deininger P: Rodent BC1 RNA gene as a master gene for ID element amplification. Proc Natl Acad Sci USA 1994; 91:3607–3611.
47 Gilbert N, Labuda D: CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs. Proc Natl Acad Sci USA 1999;96:2869–2874.
48 Santangelo AM, de Souza FS, Franchini LF, Bumaschny VF, Low MJ, Rubinstein M: Ancient exaptation of a CORE-SINE retroposon into a highly conserved mammalian neuronal enhancer of the proopiomelanocortin gene. PLoS Genet 2007; 3:1813–1826.
49 Sela N, Mersch B, Gal-Mark N, Lev-Maor G, Hotz-Wagenblatt A, Ast G: Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. Genome Biol 2007;8:R127.
50 Brosius J, Gould SJ: On 'genomenclature': a comprehensive (and respectful) taxonomy for pseudogenes and other 'junk DNA'. Proc Natl Acad Sci USA 1992;89:10706–10710.
51 Krull M, Petrusma M, Makalowski W, Brosius J, Schmitz J: Functional persistence of exonized mammalian-wide interspersed repeat elements (MIRs). Genome Res 2007;17:1139–1145.
52 Lev-Maor G, Sorek R, Shomron N, Ast G: The birth of an alternatively spliced exon: 3′ splice-site selection in Alu exons. Science 2003;300:1288–1291.
53 Gerber A, O'Connell MA, Keller W: Two forms of human double-stranded RNA-specific editase 1 (hRED1) generated by the insertion of an Alu cassette. RNA 1997;3:453–463.
54 Lev-Maor G, Sorek R, Levanon EY, Paz N, Eisenberg E, Ast G: RNA-editing-mediated exon evolution. Genome Biol 2007;8:R29.
55 Möller-Krull M, Zemann A, Roos C, Brosius J, Schmitz J: Beyond DNA: RNA editing and steps toward *Alu* exonization in primates. J Mol Biol 2008;382:601–609.
56 Sakurai M, Yano T, Kawabata H, Ueda H, Suzuki T: Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome. Nat Chem Biol 2010;6:733–740.
57 Sorek R: The birth of new exons: mechanisms and evolutionary consequences. RNA 2007;13:1603–1608.
58 Rollins RA, Haghighi F, Edwards JR, Das R, Zhang MQ, et al: Large-scale structure of genomic methylation patterns. Genome Res 2006;16:157–163.
59 Beauregard A, Curcio MJ, Belfort M: The take and give between retrotransposable elements and their hosts. Ann Rev Genet 2008;42:587–617.
60 Soifer HS, Zaragoza A, Peyvan M, Behlke MA, Rossi JJ: A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon. Nucleic Acids Res 2005;33:846–856.
61 Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, et al: A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. Cell 2007;129:69–82.

Jürgen Schmitz
Institute of Experimental Pathology (ZMBE)
University of Münster, Von-Esmarch-Str. 56
DE–48149 Münster (Germany)
Tel. +49 251 835 2133, E-Mail jueschm@uni-muenster.de