Evolutionary Emergence of Genes Through Retrotransposition

Richard Cordaux, *University of Poitiers, Poitiers, France*

Mark A Batzer, *Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana, USA*

Advanced article

Article Contents

doi: 10.1002/9780470015902.a0020783

Variation in the number of genes among species indicates that new genes are continuously generated over evolutionary times. Evidence is accumulating that transposable elements, including retrotransposons (which account for about 90% of all transposable elements inserted in primate genomes), are potent mediators of new gene origination. Retrotransposons have fostered genetic innovation during human and primate evolution through: (i) alteration of structure and/or expression of pre-existing genes following their insertion, (ii) recruitment (or domestication) of their coding sequence by the host genome and (iii) their ability to mediate gene duplication via ectopic recombination, sequence transduction and gene retrotransposition.

Introduction

Variation in the number of genes among species indicates that new genes are continuously generated over evolutionary times. Although the emergence of new genes and functions is of central importance to the evolution of species, studies on the formation of genetic innovations have only recently become possible. Recent advances have been made possible thanks to the availability of new molecular techniques and analytical methods, but they were also greatly supported by continuously increasing amounts of genomic data from multiple organisms, including primates.

Comparative studies have revealed that new genes most generally arise by rearrangements between pre-existing genetic structures such as exons, introns or regulatory regions. This notion actually predates the genomic era, as it was initially proposed by François Jacob in 1977 in its concept of 'evolutionary tinkering'. Under this view, the cell can metaphorically be seen as a tinkerer who works with no predefined plan, trying to use whatever is already available to produce innovations that might turn out to be beneficial. The notion of evolutionary tinkering is well illustrated by molecular mechanisms that have been shown to be involved in the creation of new gene structures. Indeed, one of the most important processes leading to the formation of new genes is through duplication of pre-existing genes followed by nucleotide sequence divergence. Other known mechanisms include exon shuffling mediated by ectopic recombination of exons and domains from distinct genes, lateral gene transfer between organisms, gene fusion or fission in which two neighbouring genes are fused to form one new gene or one gene is split into two separate genes, respectively, and *de novo* origination from previously noncoding genomic sequence. Genome sequencing projects have also highlighted that new gene structures can arise as a result of the activity of transposable elements (TEs), which are mobile genetic units or 'jumping genes' that have been bombarding the genomes of most species during evolution. For example, there are over three million TE copies inserted in the human genome, encompassing about half of its entire genome sequence.

TEs essentially fall into one of two categories, depending on their mode of transposition: class 1 elements or retrotransposons, which mobilize via a 'copy-and-paste' mechanism using a ribonucleic acid (RNA) intermediate and class 2 elements or deoxyribonucleic acid (DNA) transposons, which mobilize via a 'cut-and-paste' mechanism using a DNA intermediate. Retrotransposons represent about 90% of all TEs inserted in primate genomes and they can be classified into two major groups, depending whether they possess or lack long terminal repeats (LTRs) at both ends. LTR retrotransposons are retrovirus-like elements covering ~8% of the human genome and exemplified by human endogenous retrovirus (HERV) elements (**Figure 1**). Non-LTR retrotransposons are the most successful TEs present in primate genomes and they are typified by the autonomous LINE-1 (or L1) element (~17% of the human genome) and the nonautonomous *Alu* and SVA elements (~11% of the human genome) (**Figure 1**). Retrotransposons have been amplifying throughout all of primate evolution and some families are still actively retrotransposing in humans, as demonstrated by the existence of *de novo* L1, *Alu* and SVA insertions responsible for genetic disorders. While this kind of mutagenic effect spectacularly illustrates the deleterious impact TEs may exert

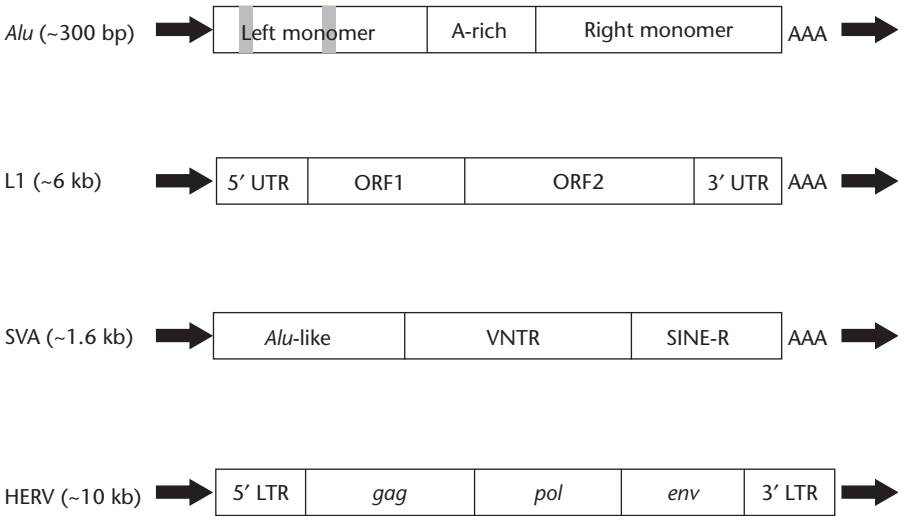

Figure 1 Structures of four typical human retrotransposons (not drawn to scale). The *Alu* element consists of two 7SL RNA-related monomers separated by an A-rich connector; the left monomer contains A and B boxes (grey boxes) promoting transcription by RNA polymerase III. The L1 element consists of two open reading frames (ORF1 and ORF2) surrounded by 5′ and 3′ untranslated regions (UTR). The SVA element consists of a region derived from a SINE-R element and an *Alu*-like region separated by a variable number of tandem repeats (VNTR). All three elements end with a poly A tail (AAA). *Alu* and SVA elements are nonautonomous retrotransposons that hijack the molecular retrotransposition machinery of the autonomous L1 element to mediate their own retrotransposition. The HERV element consists of three genes (*gag*, *pol* and *env*) surrounded by long terminal repeats (LTR). All four elements generate target site duplications (black arrows) upon insertion.

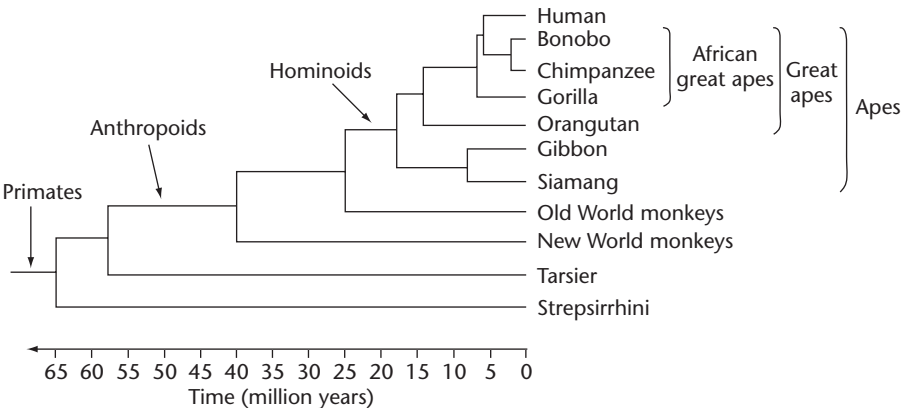

Figure 2 Schematic phylogenetic tree of the primate order. Names and approximate evolutionary age of the major lineages discussed in the main text are shown.

on the genome, continuous retrotransposon activity over evolutionary time scales sometimes has also been beneficial, for example, by promoting new gene formation. In this review, we explore mechanisms through which retrotransposons have fostered genetic innovations during human and primate evolution (**Figure 2**), including: (i) alteration of structure and/or expression of pre-existing genes following retrotransposon insertion, (ii) recruitment (or domestication) of retrotransposon coding sequence by the host genome and (iii) ability of retrotransposons to mediate gene duplication via ectopic recombination, sequence transduction and gene retrotransposition.

Gene Alteration Following Retrotransposon Insertion

Retrotransposon insertions within a gene can have different consequences, depending where in the gene the insertion occurred. It may alter gene structure or expression with a deleterious effect, as in the case of insertions responsible for genetic disorders. However, retrotransposon insertions can occasionally have a beneficial effect on the phenotype. Detecting such advantageous mutations is a difficult task, and availability of the human genome sequence has been instrumental to the recognition of TE contribution to human and primate gene evolution.

Alteration of gene structure

Retrotransposon insertions within gene-coding sequences most often result in gene inactivation. However, analysis of ~14 000 human genes indicated that ~4% contain TEs (mostly retrotransposons such as *Alu* and L1 elements) within protein-coding sequences (Nekrutenko and Li, 2001). While ~10% of the cases resulted from insertions within exons, the vast majority of cases corresponded to TEs that had inserted in introns and subsequently had been recruited as novel exons, in a process termed exonization. This high rate of TE exonization is thought to be facilitated by the fact that many TEs carry cryptic donor and acceptor splice sites, as illustrated by the *Alu* retrotransposon (**Figure 1**).

Alu elements represent the most successful TE family in the human genome, with over one million copies. The typical ~300 bp-long *Alu* nucleotide sequence contains nine guanine–thymine (GT) dinucleotides and 14 adenine–guanine (AG) dinucleotides that represent as many cryptic donor (GT) and acceptor (AG) splice sites (Sorek *et al*., 2002). Only four of these potential splice sites are located on the plus strand of the *Alu* element, whereas there are 19 on the minus strand. Thus, it is more likely that intronic *Alu* elements can be exonized when they are inserted in opposite orientation relative to host gene transcription (**Figure 3a**). Consistently, genome-wide analyses have estimated that ~5% of alternatively spliced exons are derived from *Alu* elements in humans. Because all *Alu*-containing exons are alternatively spliced, they increase transcriptome variation while preserving original protein formation. In principle, this could allow a gene containing an *Alu*-derived exon to acquire a novel function that might be beneficial to the organism.

While *Alu* elements can alter gene structure by contributing new exonic sequences to genes, they also have the potential to alter gene structure via deletion of pre-existing exons. Indeed, because *Alu* elements and other TEs represent large multigene families within the genome (e.g. there is one *Alu* element inserted every ~3 kb on average in the human genome), they may act as sources for ectopic (nonallelic) homologous recombination, i.e. recombination between homologous sequences located at nonhomologous genomic sites. Hence, ectopic recombination between two intronic *Alu* elements can result in exon deletion (**Figure 3b**). Various forms of cancers and genetic disorders have been shown to be attributable to this kind of mutation in humans. However, comparison of *Alu* recombination-mediated deletions in the human and chimpanzee genomes identified three instances in the human genome of entire exon deletions from genes that are functional in the chimpanzee genome (Sen *et al*., 2006). This indicates that ectopic recombination between TEs can generate diversity in gene structure without necessarily being deleterious. The evolutionary benefit (if any) of these three exon deletions remains unknown, but one recombination event deleted the fourth exon of *CHRNA9*, a gene with a probable role in modulation of auditory stimuli. It has been speculated that changes in *CHRNA9* gene structure resulting from ectopic recombination between *Alu* elements might be responsible for some of the unique auditory traits that distinguish humans and chimpanzees.

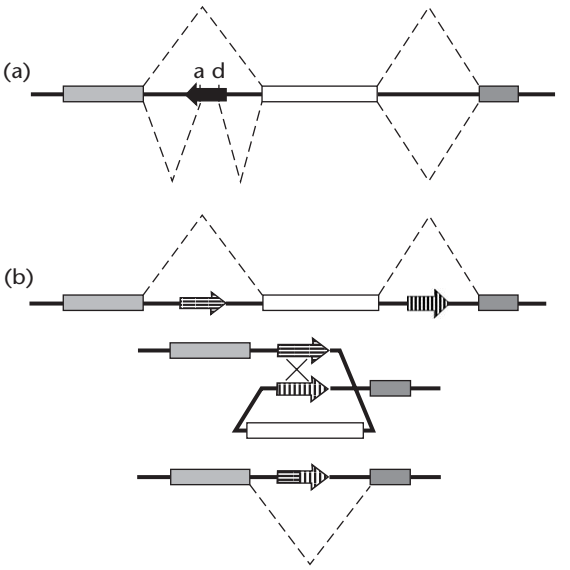

Figure 3 Alteration of gene structure mediated by *Alu* retrotransposons. (a) *Alu* exonization: a hypothetical gene constituted of three exons (light grey, white and dark grey boxes) is shown with its splicing pattern (dashed lines above gene). Activation of the cryptic donor (d) and acceptor (a) splice sites of an *Alu* element (black arrow) inserted in opposite orientation relative to gene transcription in the first intron leads to integration of noncoding *Alu* sequence in the gene's transcript (dashed lines below gene) and conversion to coding sequence. (b) Ectopic recombination: a hypothetical gene constituted of three exons (light grey, white and dark grey boxes) is shown on top with its splicing pattern (dashed lines). Ectopic recombination (crossed thin lines) between two intronic *Alu* elements (dashed arrows) leads to the deletion of the intervening sequence containing the entire white exon (middle). As a result (bottom), the gene is now constituted by two exons (light and dark grey) with a new splicing pattern (dashed lines).

The L1 retrotransposon (**Figure 1**) can also alter gene structure when inserted either in direct or opposite orientation relative to host gene transcription. Indeed, the A-rich coding strand of the full-length human L1 element contains 19 potential canonical and noncanonical polyadenylation signals that specify transcription termination (Perepelitsa-Belancio and Deininger, 2003). While this mechanism may primarily have evolved to limit L1 retrotransposition activity, an intronic L1 element inserted in the same orientation as gene transcription may cause premature termination of that gene's RNA transcripts. Similarly, the L1 nucleotide sequence possesses a 5′ antisense promoter that can drive outward transcription, as well as a 3′ antisense polyadenylation signal. Empirical evidence has recently been provided that at least three human genes had apparently been split by intronic L1 elements inserted in antisense orientation, with a transcript containing the upstream exons and terminating at the L1 antisense polyadenylation signal and a transcript derived from the L1 antisense promoter and including the downstream exons of the gene (Wheelan *et al*., 2005). These observations offer an important basis for the mechanism underlying the emergence of new gene structures by gene fission.

Alteration of gene expression

Analyses of TE prevalence within gene untranslated regions (UTR) revealed that ~27% of human genes have at least one messenger RNA (mRNA) with a TE-derived sequence within the 5′- or 3′-UTR. TE insertion upstream of a transcription unit can create a new promoter that may induce different gene expression patterns for example between species, between tissues or between different members of gene families. The *CYP19* gene provides an example in which a retrotransposon sequence is used as a promoter to drive tissue-specific expression in anthropoid primates (van de Lagemaat *et al.*, 2003).

The *CYP19* gene encodes aromatase P450, a key enzyme involved in oestrogen synthesis. This gene is expressed in gonads and brain of most mammals. Moreover, in primates, *CYP19* is also expressed at high levels in placenta. Expression in different tissues is linked to the presence of alternative promoters. In primates, another alternative promoter located ~100 kb upstream of the coding region drives placental-specific expression. This promoter was shown to correspond to the LTR sequence of an HERV retrotransposon (**Figure 1**), which is inserted at the orthologous locus in hominoids, Old World and New World monkeys. This suggests that this LTR sequence was acquired by an anthropoid ancestor and used as an alternative promoter controlling oestrogen levels during pregnancy.

Retrotransposon insertion upstream of a gene can also modify regulation as to increase or decrease gene expression. For example, the *apo(a)* gene, which encodes one of the components of plasma lipoprotein particles in human, presents an allelic polymorphism in which one of the alleles possesses a *cis*-activating sequence that leads to a 10-fold increase in promoter activity in cell culture assays, compared to the most frequent allele. This regulatory sequence is part of the 5′ region of an L1 retrotransposon (**Figure 1**) inserted ~2 kb upstream of the transcription initiation site of the *apo(a)* gene.

While there is now ample evidence for the use of the sequence of particular retrotransposon copies for the regulation of gene expression at specific loci, a recently but increasingly recognized phenomenon is the recruitment of nonautonomous retrotransposon families as functional noncoding elements. Typically, these are TE families in which tens or hundreds of copies preferentially located near genes appear to exhibit conserved nucleotide sequences over long evolutionary times, suggesting that retrotransposons may have massively impacted regulation of gene expression in humans and other organisms during evolution.

Retrotransposon Recruitment by Host Genome

The process of TE recruitment by the genome, also known as molecular domestication, is the use by the organism of a function carried by a TE. TEs have multiple characteristics that predispose them to be recruited to fulfil cellular functions, as they encode proteins that can for example bind, copy, break, join or degrade nucleic acids. TE sequences have been repeatedly domesticated during eukaryotic evolution, and many candidates have been revealed by genome sequencing projects. Hence, by conservative estimates, about 50 genes derived from TEs were identified in the human genome. The vast majority of these genes appear to be derived from DNA transposons. A typical such example is provided by the RAG1 protein involved in V(D)J recombination, a key step in the evolution of adaptive immune system in jawed vertebrates, which is derived from *Transib* DNA transposons. There also exist cases of retrotransposons that have been recruited by their host, such as the *TART* and *Het-A* non-LTR retrotransposons that maintain the telomeres of linear chromosomes in *Drosophila* flies. Contrary to most organisms whose telomeres consist of arrays of simple repeats generated by a telomerase, *Drosophila* lacks telomerase and its telomeres consist of arrays of *TART* and *Het-A* elements generated by retrotransposition.

Retrotransposons have also been recruited by the genome during primate evolution, as illustrated by the *syncytin-1* and *syncytin-2* genes. These genes are involved in placental formation and have been shown to be derived from *env* genes of HERV LTR retrotransposons (**Figure 1**). The ~30 000 HERV copies inserted in the human genome are proviral remnants of ancestral germline infections by active retroviruses, which have been inactivated by nucleotide mutations and/or deletions and are now vertically inherited. However, computational searches of the human genome for retroviral genes from 39 HERV families identified ~500 complete *env* genes, 16 of which encoded intact and therefore potentially functional proteins (Blaise *et al.*, 2003). The *env* gene encodes the viral envelope glycoprotein that normally mediates virus entry into target cell through binding to a cellular receptor and virus/cell fusion. Functional assays indicated that 2 (*syncytin-1* and *syncytin-2*) of the 16 *env* genes had preserved fusogenic activities.

Notably, *syncytin-1* had previously been shown to be derived from a HERV-W *env* gene (Mi *et al.*, 2000). Examination of *syncitin-1* gene expression in 23 human tissues by northern blot revealed strong expression in placenta and weak to no expression in all other tissues examined. Further analyses of placental tissues indicated that *syncitin-1* expression is restricted to placental syncytiotrophoblasts, which are multinucleated cells that originate from fusion of fetal trophoblasts. These cells constitute the boundary layer between maternal and fetal tissue. They are important in maternal–fetal exchange, in tissue remodelling during placental development and in protecting the developing fetus from the maternal immune response. Results of *in vitro* culture experiments were also consistent with the implication of *syncytin-1* in fetal trophoblast fusion and placental syncytiotrophoblast layer formation.

Comparative analyses of the sequence homologous to chromosomal region 7q21.2 where *syncytin-1* is located in humans have shown that an HERV-W retrotransposon inserted in the primate genome 25–40 million years ago, as

it is present in Old World monkeys and hominoids, but absent in New World monkeys. *Syncytin-1* is subject to negative selection and conserved in hominoid primates. By contrast, the orthologous *syncytin-1* locus in Old World monkeys has experienced deletions and nucleotide mutations introducing frameshifts and premature stop codons, thus rendering the locus nonfunctional. This indicates that the recruitment of the HERV-W retrotransposon took place 18–25 million years ago, specifically in hominoids, the primate lineage leading to humans.

Retrotransposons have been involved in domestication processes not only by directly contributing their coding sequences and functions, but also indirectly via subtle contributions facilitating the domestication of other transposable elements, as exemplified by the *SETMAR* gene. *SETMAR* is a primate chimaeric gene resulting from fusion of a *SET* histone methyltransferase gene to the transposase gene of an *Hsmar1* DNA transposon. Comparative analyses showed that the recruitment of the transposase gene as part of *SETMAR* took place in an anthropoid primate ancestor 40–58 million years ago, after the insertion of an *Hsmar1* transposon downstream of a pre-existing *SET* gene, followed by *de novo* exonization of previously noncoding sequence and the creation of a new intron (Cordaux *et al*., 2006). However, the birth of this gene might have never occurred without the contribution of a retrotransposon. Indeed, all anthropoid species carrying the *Hsmar1* transposon also share an *Alu*Sx retrotransposon (**Figure 1**) inserted in the 5′-terminal inverted repeat of the *Hsmar1* element, which was partially deleted during the *Alu*Sx integration process. Because both terminal inverted repeats of transposons are necessary for transposition, the *Alu*Sx insertion may have contributed to the recruitment of the *Hsmar1* transposon as part of *SETMAR* by immobilizing it at a time when the *Hsmar1* family was experiencing high levels of transposition in primate genomes.

Retrotransposon-mediated Gene Duplication

As outlined in the introduction, formation of genes through duplication of pre-existing genes followed by nucleotide sequence divergence is an important source of new genes. This mechanism is responsible for the origin of numerous gene families, such as the opsin gene family involved in colour vision in primates. In New World monkeys, colour vision depends on two opsin proteins that are encoded by two genes, one absorbing blue wavelengths and the other one absorbing red and green wavelengths. As a result, New World monkeys have a dichromatic vision, i.e. they can see blue objects but cannot distinguish between red and green. By contrast, in the primate lineage leading to Old World monkeys and hominoids, the red/green opsin gene was duplicated about 25–40 million years ago. The two copies diverged from each other and became specialized so that one gene now encodes an opsin protein that absorbs red wavelengths while the opsin protein encoded by the other gene absorbs green wavelengths. These primate species thus have evolved trichromatic vision, i.e. they can distinguish blue, red and green colours. TEs represent potent sources of duplication. For example, retrotransposons can mediate gene duplication as direct consequences of their insertion and involvement in ectopic recombination, or as byproducts of the retrotransposition mechanism that mediates their mobility (sequence transduction and gene retrotransposition).

Ectopic recombination

Ectopic recombination between dispersed TEs can result in various genomic rearrangements, such as duplications, deletions and translocations. Interestingly, compared to other sequenced organisms, the human genome is enriched in large (>10 kb in length) and highly homologous ($>90\%$ sequence identity) duplicated genomic regions, termed segmental duplications. Detailed analyses of human segmental duplication boundaries identified a significant enrichment in *Alu* retrotransposons (**Figure 1**), in that *Alu* elements comprise $\sim$24% of boundary sequences, compared with $\sim$11% elsewhere in the human genome (Bailey *et al*., 2003). Considering that $\sim$5% of the human genome has been duplicated within the past $\sim$40 million years, it has been proposed that recombination between *Alu* inserts may represent an important mechanism for the origin and expansion of segmental duplications during primate evolution.

Because segmental duplications encompass large genomic regions, they represent a potent pathway for duplicating entire genes. This is well illustrated by the emergence of the *morpheus* gene family during hominoid evolution. Analysis of the human genome sequence identified a $\sim$20 kb-long DNA segment, termed LCR16a, duplicated 15 times throughout $\sim$15 Mb of the short arm of human chromosome 16 (Johnson *et al*., 2001). Further genomic sequence analyses and fluorescence *in situ* hybridization experiments revealed that LCR16a proliferation occurred during the past 25 million years of hominoid evolution, since Old World monkeys carry only one or two LCR16a copies. By contrast, great apes and human carry 9–30 LCR16a copies, some of which are shared between species, and others specific to particular lineages. Interestingly, LCR16a was found to contain a gene that was duplicated as part of LCR16a, leading to the creation of the *morpheus* gene family. Alignment of human paralogous *morpheus* genes revealed 10% nucleotide divergence in coding exonic regions and only $\sim$2% divergence in noncoding intronic regions, which is indicative of positive selection and a hallmark of genes undergoing adaptive evolution. Although the precise function of the *morpheus* gene family is unknown, the fact that it is under positive selection suggests that it might fulfil an important function in humans and other apes.

Sequence transduction

Retrotransposon-mediated ectopic recombination results from the physical occurrence of retrotransposon insertions at particular sites in the genome. Therefore, it can be seen as a post-insertional mechanism of retrotransposon-mediated gene duplication. Retrotransposons can also mediate gene duplication at a pre-insertional stage of their life cycle, i.e. during the retrotransposition process *per se*. For example, during L1 retrotransposition, an L1 genomic copy (**Figure 1**) is transcribed into RNA by RNA polymerase II and then exported to the cytoplasm. Translation of the L1 RNA allows the retrotransposition molecular machinery to be assembled. The L1 RNA and retrotransposition machinery are then imported back to the nucleus. After reverse transcription of the L1 RNA and integration, a novel L1 copy ends up inserted at a different genomic locus from the one that yielded the initial RNA transcript. It is noteworthy that retrotransposons usually do not carry downstream motifs allowing efficient transcription termination. Therefore, the RNA transcription machinery sometimes skips the retrotransposon own weak polyadenylation signal and terminates transcription by using an alternative polyadenylation signal located downstream in the 3′ flanking genomic sequence (**Figure 4**). The read-through transcript containing the retrotransposon along with the extra genomic sequence is subsequently integrated back into the genome via retrotransposition. This process, in which a retrotransposon carries – and therefore duplicates – 3′ flanking sequence during its own retrotransposition, is known as 3′ transduction.

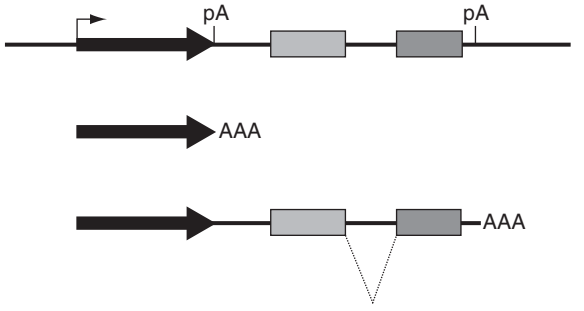

Figure 4 Retrotransposon-mediated sequence transduction. A hypothetical gene constituted of two exons (light and dark grey boxes) and an upstream L1 retrotransposon (black arrow) are shown on top with their respective polyadenylation motifs (pA) signalling transcription termination. RNA transcription starts at the 5′ end of the L1 element (thin horizontal arrow) and normally proceeds down to the L1 polyadenylation signal, resulting in transcription termination. The transcript (middle) therefore consists of the L1 RNA sequence ending with a poly A tail (AAA), which can subsequently be integrated into the genome by retrotransposition. Sometimes, the L1 polyadenylation signal is ignored and transcription proceeds down to another polyadenylation signal located in the L1 flanking sequence. The transcript therefore consists of the L1 RNA sequence, followed by the downstream sequence flanking the L1 element and a poly A tail (bottom). In this example, the downstream sequence contains a gene which intron is being spliced out (dashed lines) before transcript integration into the genome by retrotransposition, resulting in the duplication of the L1 element and an intronless version of the original gene.

In principle, 3′ transduction could lead to the duplication of coding sequences located in the transduced flanking genomic sequence. The potential of L1 retrotransposons to mediate exon shuffling via 3′ transduction has been experimentally confirmed using cell culture assays (Moran *et al.*, 1999). However, the evolutionary contribution of L1-mediated 3′ transduction to genetic innovation was supported by only a couple of putative cases of exon transduction reported in the human genome. A recent analysis of SVA retrotransposons has demonstrated the evolutionary significance of retrotransposon-mediated 3′ transduction, by showing that SVA-mediated transduction is responsible for the creation of the *AMAC1* gene family that comprises four copies in the human genome (Xing *et al.*, 2006).

Similar to L1 retrotransposons, SVA elements (**Figure 1**) can transduce 3′ flanking sequence. A genome-wide analysis indicated that ~10% of human SVA retrotransposons have been involved in the transduction of sequences ranging in size from a few dozens of nucleotides to almost 2 kb. Interestingly, three transduced sequences located on chromosomes 8, 17 and 18 were found to originate from the same source locus located elsewhere on chromosome 17. Analysis of the four paralogous sequences identified four copies of the *AMAC1* gene. The ancestral *AMAC1L3* copy at the source locus consisted of two exons separated by an intron. By contrast, the three transduced copies were intronless versions of *AMAC1L3*, as a result of the splicing of the intron during the retrotransposition process (**Figure 4**). Evolutionary analyses demonstrated that the three transduction events all took place 7–14 million years ago, as human and African great apes share all four *AMAC1* copies, whereas orangutan possesses only the ancestral *AMAC1L3* gene. Experimental studies indicated that, in addition to *AMAC1L3*, at least two of the three transduced *AMAC1* genes are expressed in human tissues. RNA transcript sequence analyses of the expressed *AMAC1* duplicates further revealed that the promoter sequence had been duplicated along with the *AMAC1* coding sequence as part of the 3′ transduction process. This is an interesting observation because it indicates that retrotransposon-mediated gene transduction cannot only duplicate coding regions of genes but also their regulatory regions, thus retaining functional potential after duplication. Hence, this retrotransposon-mediated duplication mechanism can lead to rapid generation of functional gene families.

Gene retrotransposition

Gene retrotransposition is another mechanism that can duplicate genes. Similar to 3′ transduction, gene retrotransposition operates during the retrotransposition process itself. However, contrary to 3′ transduction, gene retrotransposition only duplicates gene sequences and no retrotransposon sequence is coduplicated in the process. This is because gene retrotransposition is based on the 'hijacking' of the L1 retrotransposition molecular machinery by host gene mRNA transcripts, which are reverse

transcribed and reintegrated in the genome. As a result, similar to 3′ transduction, this mechanism generates intronless duplicate copies of the original genes. However, in sharp contrast with 3′ transduction, gene retrotransposition generally does not duplicate upstream regulatory regions, thus requiring duplicated genes to fortuitously acquire new regulatory regions to be functional. As a consequence, gene retrotransposition has long been thought to generate so-called retropseudogenes, i.e. nonfunctional gene copies rapidly accumulating deletions and nucleotide mutations introducing frameshifts and premature stop codons. However, recent studies have shown that gene retrotransposition has generated a number of new functional genes (called retrogenes), as illustrated by the emergence of the *GLUD2* retrogene during hominoid evolution (Burki and Kaessmann, 2004).

The enzyme glutamate deshydrogenase is important for recycling the excitatory neurotransmitter glutamate, during neurotransmission. In human, this enzyme exists as two isotypes encoded by the *GLUD1* and *GLUD2* genes. *GLUD1* is a housekeeping gene expressed in many tissues whereas *GLUD2* is specifically expressed in nerve tissues and testis. *GLUD1* is present in the genome of primates and nonprimate mammals such as mouse and rat. By contrast, *GLUD2* is specific to hominoid primates, as it is present in human and apes but lacking at the orthologous locus in Old World monkeys. Evolutionary analyses indicated that the X-linked intronless *GLUD2* gene originated 18–25 million years ago in a hominoid ancestor, through retrotransposition of a spliced mRNA from the intron-containing *GLUD1* gene located on human chromosome 10. Sequence analyses further revealed that *GLUD1* was subject to strong purifying selection after gene duplication and maintained its original housekeeping function. By contrast, *GLUD2* early evolution was driven by positive selection, indicative of functional adaptation, followed by purifying selection maintaining the newly arisen functional variants.

Genome-wide searches have confirmed the importance of retrotransposition in the emergence of new primate genes. It has been estimated that at least one new retrogene per million year emerged on the human lineage during the past $\sim$65 million years of primate evolution (Marques *et al*., 2005). Interestingly, the majority of these retrogenes are specifically or predominantly expressed in testis whereas their source genes are expressed ubiquitously. Consistently, most retrogenes have evolved functional roles in spermatogenesis. Therefore, emerging retrogenes may have been recruited during primate evolution by natural and/or sexual selection to enhance male germline function.

Conclusion

TEs have long been considered as junk DNA. However, genome sequencing projects have highlighted the high proportion of TEs present in many genomes and the many effects by which TEs have impacted and continue to impact the genomes where they reside. TE involvement in genetic innovation is one of the many ways they have been shaping genome evolution. Does their beneficial influence on host genome explain the evolutionary success of so many TE families? Probably not because in most cases, at best small subsets of all copies from given TE families turn out to have benefited the genome. Whatever function or 'usefulness' can be hypothesized to explain TE colonization and maintenance in genomes over long evolutionary time scales, it is likely that TEs mostly owe their success to their own capacity to spread in their host genomes. Therefore, TE contribution to shaping genome evolution appears to be a consequence, not a reason, for their maintenance in host genomes.

References

Bailey JA, Liu G and Eichler EE (2003) An *Alu* transposition model for the origin and expansion of human segmental duplications. *American Journal of Human Genetics* **73**: 823–834.

Blaise S, de Parseval N, Bénit L and Heidmann T (2003) Genome-wide screening for fusogenic human endogenous retrovirus envelopes identifies *syncytin 2*, a gene conserved on primate evolution. *Proceedings of the National Academy of Sciences of the USA* **100**: 13013–13018.

Burki F and Kaessmann H (2004) Birth and adaptive evolution of a hominoid gene that supports high neurotransmitter flux. *Nature Genetics* **36**: 1061–1063.

Cordaux R, Udit S, Batzer MA and Feschotte C (2006) Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proceedings of the National Academy of Sciences of the USA* **103**: 8101–8106.

Johnson ME, Viggiano L, Bailey JA *et al*. (2001) Positive selection of a gene family during the emergence of humans and African apes. *Nature* **413**: 514–519.

van de Lagemaat LN, Landry JR, Mager DL and Medstrand P (2003) Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends in Genetics* **19**: 530–536.

Marques AC, Dupanloup I, Vinckenbosch N, Reymond A and Kaessmann H (2005) Emergence of young human genes after a burst of retroposition in primates. *PLoS Biology* **3**: e357.

Mi S, Lee X, Li X *et al*. (2000) Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**: 785–789.

Moran JV, DeBerardinis RJ and Kazazian HH Jr (1999) Exon shuffling by L1 retrotransposition. *Science* **283**: 1530–1534.

Nekrutenko A and Li WH (2001) Transposable elements are found in a large number of human protein-coding genes. *Trends in Genetics* **17**: 619–621.

Perepelitsa-Belancio V and Deininger P (2003) RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nature Genetics* **35**: 363–366.

Sen SK, Han K, Wang J *et al*. (2006) Human genomic deletions mediated by recombination between *Alu* elements. *American Journal of Human Genetics* **79**: 41–53.

Sorek R, Ast G and Graur D (2002) Alu-containing exons are alternatively spliced. *Genome Research* **12**: 1060–1067.

Wheelan SJ, Aizawa Y, Han JS and Boeke JD (2005) Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. *Genome Research* **15**: 1073–1078.

Xing J, Wang H, Belancio VP *et al*. (2006) Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proceedings of the National Academy of Sciences of the USA* **103**: 17608–17613.

Further Reading

Batzer MA and Deininger PL (2002) Alu repeats and human genomic diversity. *Nature Reviews Genetics* **3**: 370–379.

Chen JM, Stenson PD, Cooper DN and Ferec C (2005) A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. *Human Genetics* **117**: 411–427.

Craig NL, Craigie R, Gellert M and Lambowitz AM (2002) *Mobile DNA II*. Washington, DC: American Society for Microbiology Press.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.

Jacob F (1977) Evolution and tinkering. *Science* **196**: 1161–1166.

Long M, Betran E, Thornton K and Wang W (2003) The origin of new genes: glimpses from the young and old. *Nature Reviews Genetics* **4**: 865–875.

Ostertag EM and Kazazian HH Jr (2001) Biology of mammalian L1 retrotransposons. *Annual Reviews of Genetics* **35**: 501–538.

de Parseval N and Heidmann T (2005) Human endogenous retroviruses: from infectious elements to human genes. *Cytogenetic and Genome Research* **110**: 318–332.

Volff JN (2006) Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *Bioessays* **28**: 913–922.

Wang H, Xing J, Grover D *et al*. (2005) SVA elements: a hominid specific retroposon family. *Journal of Molecular Biology* **354**: 994–1007.