Evolution of Human Retrosequences: Alu

Astrid M Roy-Engel, Tulane University Health Sciences Center, New Orleans, Louisiana, USA
Mark A Batzer, Louisiana State University, Baton Rouge, Louisiana, USA
Prescott L Deininger, Tulane University Health Sciences Center, New Orleans, Louisiana, USA

Throughout evolution, mobile elements have accumulated to high copy numbers contributing to almost half of the human genomic mass. Evidence indicates that only the retroelements are currently active. In humans, the short interspersed elements (SINE), Alu, with over one million copies, outnumbers any of the other types of retroelements. Alu arose from the dimerization of modified 7SL RNA (ribonucleic acid) pseudogenes early in primate evolution, where different subfamilies continued to amplify during particular periods. Alu amplification has both positively and negatively impacted the human genome, and continues to play an important role in its shaping as a contributor of genetic instability and variation.

Retroelement Features and Classification

A retroelement is defined as an element that amplifies itself in the genome via a ribonucleic acid (RNA) intermediate, but does not have an infectious form. Several viruses like retroviruses, caulimoviruses and hepadnaviruses use reverse transcription of RNA as a step in their cycle, but are not considered retroelements. However, some of the viruses are related to retroelements. Retroviruses are thought to have evolved from LTR (long-terminal repeat) retrotransposons (Malik et al., 2000), as they share several features in common (Figure 1). There are several types of retroelements: LTR retrotransposons, non-LTR retrotransposons (long interspersed elements, LINE-like elements), nonautonomous retroelements (the short interspersed elements, SINEs; SINE-RVNTR-Alu element, SVA) and retropseudogenes (Weiner et al., 1986; see Figure 1). Some of the retroelements code for the enzymes and proteins they require for amplification. In contrast, the nonautonomous elements (SINEs, SVA and retropseudogenes) need to ‘parasitize’ the factors they require from external sources. See also: Retroviral Repeat Sequences

Human Mobile Retroelements

The human genome contains several types of mobile elements. Among them are the retroelements, which account for about 40% of its mass, with a substantial part being LINEs and SINEs (Lander et al., 2001). Retroelements are classified into two groups: LTR and non-LTR retrotransposons. LTR retrotransposons make up 8% of the human genome, including MaLR, Mer4 and several families of endogenous retroviruses like HERV-K elements. The human genome contains two families of non-LTR retrotransposons: an older inactive LINE family L2 that accounts for about 3–4% of the human genome and the currently active L1 accounting for about 16%. Most of the nonautonomous, non-LTR retrotransposons fall into a general class, termed SINEs that derive from polymerase III (pol III) transcripts. Other nonautonomous elements include SVA and retropseudogenes or processed pseudogenes (Chen et al., 2005). Estimates indicate that the human genome contains from 23,000 to 33,000 retropseudogenes (Goncalves et al., 2000) and around 2,700 SVA elements (Wang et al., 2005). Although there are different types of SINEs (like MIRs), Alu elements are the most abundant in the human genome. Over one million copies of Alu elements are present in the human genome, contributing to approximately 10% of its mass. See also: Short Interspersed Elements (SINEs)

Origin of Alu Elements

Alu elements are dimeric molecules, composed of two nonidentical units or arms joined in the middle by an A-rich
region (Weiner et al., 1986). Two phases are observed in Alu evolution: an ancient monomeric period with the origins of the progenitor sequences leading to the Alu family (Figure 2), and the evolutionarily more recent period (discussed later) involving the amplification of dimeric sequences (Figure 3). Alu elements are proposed to have originated from a partial deletion of a pseudogene of the 7SL RNA gene, an integral part of the signal recognition particle (SRP) involved in protein secretion (Weiner et al., 1986). It is suggested that an Alu arm arose through the deletion of the central 60% of an approximately 300-bp long 7SL molecule. This first old Alu-like monomorphic family or fossil Alu monomer (FAM) possibly arose early in the mammalian radiation (Labuda and Zietkiewicz, 1994). Because the FAM still contains the internal pol III promoter of the 7SL, it has the potential to transcribe RNA and amplify. Amplification and subsequent evolution of the FAM family generated the free left Alu (FLA) monomer and the free right Alu (FRA) monomer families with sequence variations between each other. The first progenitor of the Alu dimeric family possibly arose through the fusion of FLA and FRA monomers. The left monomer contains the internal pol III promoter with the A and B boxes. The Alu RNA contains monomers are separated by an A-rich region and the 3' flank contains a poly A tract.

**Figure 2** Origin of Alu elements. Alu elements are thought to have arisen from a processed 7SL RNA giving rise to the ancestral element: fossil Alu monomer (FAM). FAM evolved to the FLA monomer and the free right Alu (FRA) monomer families with sequence variations between each other. The first progenitor of the Alu dimeric family possibly arose through the fusion of FLA and FRA monomers. The left monomer contains the internal pol III promoter with the A and B boxes. The Alu RNA contains monomers are separated by an A-rich region and the 3' flank contains a poly A tract.

**Figure 3** Evolutionary tree of the Alu subfamilies and amplification rates throughout the primate radiation. The old Alu subfamilies J, Sx and Sg1 were most active around 35–55 mya (indicated at the right) giving rise to the majority of the Alu elements present today in the human genome. Boxed areas represent the potential period of maximum activity for each Alu subfamily. The amplification rate of Alu decreased with evolutionary time as observed by the reduction of the copy numbers (indicated at the left). Currently the young Alu subfamilies (Y, Ya5, Yb8, Ya5, Ya5a2 and Yc1) contribute to all the known polymorphisms in the human genome.
C1) and one subfamily of FRA elements are currently present at modest copy numbers in the human genome. Detection of human-specific subfamilies suggests that both Alu monomer progenitor sequences were active after the emergence of humans. It is thought that the fusion of FLA and FRA monomers gave rise to the progenitor of the Alu family over 65 million years ago (mya).

Alu Elements Depend on L1 Factors for their Amplification

The evolution of mobile elements, particularly nonautonomous elements such as Alu, is intricately tied to their mechanism of evolution. Several lines of evidence and experimental data demonstrated that both Alu elements and processed pseudogenes depend on the endonuclease and reverse transcriptase activities provided in trans by the L1 elements (Esnault et al., 2000; Dewannieux et al., 2003). In a tissue culture system, supplementation with the L1 ORF2 protein is sufficient for Alu retroposition. Evolutionary, L1 preceded the emergence of Alu. Genome analyses also indicate the presence of earlier SINE (MIR) and LINE (L2) families with a similar parasitic relationship, before the primate expansion (Smit, 1999). In this case, the extinction of L2 and MIR coincided. Thus the SINE dependence on LINE factors has clearly dictated an interrelated amplification between the families of these retroelements. See also: Long Interspersed Nuclear Elements (LINEs): Evolution

Evolutionary Expansion of Alu Elements

Alu elements appear to be present only in primates, although there are some related elements in other organisms. Alu started to amplify 65 mya with peak amplification between 60 and 35 mya (Figure 3). The peak of amplification involved the formation of the J and Sx subfamilies. The current rate of amplification of approximately one Alu insertion in every 20–200 births is about two orders of magnitude slower than that peak. An accepted model of Alu amplification proposes that only a few retroposition competent or ‘master/source’ Alu elements generated the majority of copies present in the human genome (Deininger et al., 1992). Because the copies generated by this source gene are identical, any mutations in the source element would be reflected in the copies, thus creating subfamilies of Alu elements. Throughout the primate radiation, several different families of Alu elements have arisen (Figure 3). The oldest Alu subfamilies J, Sx and Sg1 amplified very efficiently early in the primate radiation (60–35 mya) contributing to >80% of all the Alu elements present in the human genome. Subsequently, the closely related ‘young’ Alu subfamilies, such as Y, Yc1, Ya5, Ya5a2, Ya8, Yb8, etc., appeared. Currently, these subfamilies are concomitantly active in humans. These subfamilies account for almost all of the recently integrated Alu elements within the human genome, creating genomic variations between different individuals known as polymorphisms. Most of these subfamilies of young Alu elements inserted after the human radiation from the African apes, therefore are not found in nonhuman primates. The young subfamilies are the only ones known to be currently active in the human genome (Batzer and Deininger, 2002).

Evolutionary Impact of Alu Elements on the Human Genome

There are over one million copies of Alu elements spread across the entire human genome (Lander et al., 2001). The amplification of Alu elements has had both a positive and negative impact. Alu has had a negative effect on the genome both through insertional mutagenesis and Alu–Alu recombination. These two mechanisms account for 0.1 and 0.3%, respectively, of the human genetic disorders (Deininger and Batzer, 1999). Alu–Alu recombination has caused disease both at the germ-line and somatic (causing cancer) levels. Some of these unequal homologous recombinations occur in a higher frequency in specific genes, such as the LDLR gene and ALL-1 gene for acute myelogenous leukemia (Deininger and Batzer, 1999). This type of recombination also contributes to the loss of genetic material (Sen et al., 2006). See also: Repetitive Elements and Human Disorders

Alu elements have also positively impacted the human genome. Recombination between Alu elements can result in deletion or duplication of sequences that may allow for duplication of exons and formation of new protein variants. In addition, Alu elements can introduce polyadenylation signals, alternate splice sites, which can lead to exonization with great effects on the human transcriptome (Sela et al., 2007). Some of the recombinations occurring between different chromosomes may have lead to alterations involved in speciation. Alu elements contain multiple CpG sites that can be methylated. Because methylation is associated with gene regulation and imprinting, an insertion of an Alu would introduce a CpG island in a new location possibly contributing to the evolution of gene expression and imprinting (Schmid, 1996). Alu and L1 elements are probably the main source of CpG islands in the genome (Kang et al., 2006). In addition, Alu elements have several half sites for steroid hormone receptors and may have contributed to the distribution of these response elements throughout the genome (Laperriere et al., 2007). There are some cases where an insertion of these sites through an Alu near a gene has altered gene expression. Overall, Alu elements have played an important role in the shaping of the human genome to what it is today, and continue to play a major role in human genetic instability. See also: Gross Insertions and Microinsertions in Evolution

References


Further Reading


