Minireview

Alu sequences

A.J. Mighell[a,b,*], A.F. Markham[a], P.A. Robinson[a,b]

[a]*Molecular Medicine Unit, The University of Leeds, Clinical Sciences Building, St. James's University Hospital, Leeds LS9 7TF, UK*
[b]*Leeds Dental Institute, Clarendon Way, Leeds LS2 9LU, UK*

Received 12 August 1997; revised version received 25 September 1997

Abstract ***Alu*** **sequences are frequently encountered during study of human genomic nucleic acid and form a major component of repetitive DNA. This review describes the origin of** ***Alu*** **sequences and their subsequent amplification and evolution into distinct subfamilies. In recent years a number of different functional roles for** ***Alu*** **sequences have been described. The multiple influences of** ***Alu*** **sequences on RNA polymerase II-mediated gene expression and the presence of** ***Alu*** **sequences in RNA polymerase III-generated transcripts are discussed.**

Key words: *Alu*; Repetitive DNA; Retrotransposon; Polymerase II; Polymerase III; Human genome

1. Alu sequences are repetitive DNA

Repetitive DNA accounts for at least 20% of the human genome, and has been classified into several different types that include four principal families of interspersed repeats; *Alu*, *Line 1*, *MIR* and *MaLR* [1]. An estimated $0.5–1.1 \times 10^6$ *Alu* sequences represent around 6–13% of human genomic DNA. They were named after the *Alu*I restriction enzyme site within the consensus *Alu* sequence [2].

Consensus *Alu* sequences are approximately 280 bp in length, and consist of two similar, but distinct monomers linked by an oligo-d(A) tract (Fig. 1). The right *Alu* monomer contains a 31 bp insert absent from the left monomer. A functional two box (A and B) RNA polymerase (pol III) promoter (type 2) is present in the left monomer, but is absent from the right monomer [3,4]. *Alu* sequences do not themselves include a $d(T)_4$ RNA pol III terminator, although one is often present in the flanking downstream genomic sequence. At the designated 3′ end of the *Alu* transcript there is a oligo-d(A) of variable length. Newly retrotransposed *Alu* sequences are flanked by direct repeats of host sequence, which is consistent with insertion into staggered nicks in the DNA [5,6]. *Alu* sequences are restricted to humans and other primates [7,8].

2. The origins of *Alu* sequences

Alu sequences are postulated to be retrotransposons that have inserted into the human genome via a single-stranded RNA intermediate generated by RNA pol III transcription [9]. The mechanisms and factors controlling retrotransposition are very poorly understood [10–13]. Structural homology exists between *Alu* sequences and human 7SL RNA, an abundant cytoplasmic moiety that is conserved among eukaryotic taxa [14]. The 300 bp 7SL RNA is an essential component of the signal recognition particle (SRP), which mediates the translocation of secretory proteins across endoplasmic reticulum.

A model for formation of the original *Alu* sequence has been proposed on the basis of statistical analysis of nucleotide variation at specific sites, and phylogenetic studies (Fig. 1). The fossil *Alu* monomer (FAM) was formed by deletion of the central 'S-domain' from 7SL RNA and addition of a 3′-d(A) tract which may have facilitated reverse transcription of RNA pol III transcripts. FAMs remain in low copy number in the human genome [15]. The free left arm monomer (FLAM) arose by deletion of 42 bp from FAM, and subsequently developed into subfamily A, and the evolutionarily younger subfamily C. BC200 RNA, a conserved cytoplasmic RNA pol III transcript of unknown function restricted to neural cells, also developed from FLAM [8,16]. Of lower abundance than FLAM within the human genome are copies

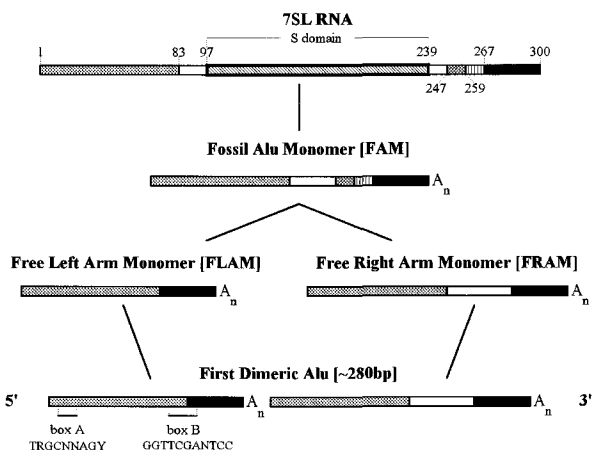

Fig. 1. Proposed model of dimeric *Alu* formation via intermediate monomeric units derived from 7SL RNA which is neither capped nor polyadenylated. Important nucleotide positions are marked on the schematic 7SL RNA moiety. See text for explanation of the progression from 7SL RNA to the first dimeric *Alu* repeat. The approximate positions and consensus sequences of the RNA pol III promoter boxes A and B in the *Alu* are marked.

*Corresponding author. Fax: (44) (113) 244 4475.
E-mail: A.J.Mighell@leeds.ac.uk

Abbreviations: FAM, fossil *Alu* monomer; fl*Alu* RNA, full-length *Alu* RNA; FLAM, free left arm monomer; FRAM, free right arm monomer; RNA pol II, RNA polymerase II; RNA pol III, RNA polymerase III; sc*Alu* RNA, small cytoplasmic *Alu* RNA; SRP, signal recognition particle

PII S0014-5793(97)01259-3

of the free right arm monomer (FRAM) which arose by deletion of 11 bp from FAM. The first *Alu* sequence was apparently formed by dimerisation of a subfamily C FLAM and a FRAM, although the mechanism by which this occurred is unclear [8,15,17].

3. Evolution of *Alu* sequences

Subsequent to formation of the first *Alu* repeat there has been a massive, but time-restricted amplification of *Alu* retrotransposons within the human genome [10]. This contrasts to a limited amplification of both their monomeric and 7SL RNA precursors. The conservation of 7SL RNA across all species suggests that monomeric precursors of *Alu* sequences were present in all lineages.

Statistical analysis has identified key diagnostic nucleotide positions in *Alu* sequences that define 12 subfamilies. A consensus nomenclature for *Alu* subfamilies has been agreed [18] (Fig. 2). As more sequences become available for analysis it is possible that further *Alu* subfamilies may be identified. Phylogenetic studies of orthologous primate loci indicate that *Alu* subfamilies were sequentially inserted into the genome. Together with computer modelling this has allowed estimations of *Alu* subfamily age [19–21] (Fig. 2).

Amplification of *Alu* subfamilies has not occurred at a constant rate throughout evolution. The majority of *Alu* retrotranspositions were completed at least 30 million years ago when the *Alu-Sx* subfamily, which accounts for half of all human *Alu* sequences, and the *Alu-Sp* and *Alu-Sq* subfamilies became unable to replicate [20]. *Alu* retrotransposition still continues in the youngest subfamilies, but is a rare event [22]. There are only around 1500 *Alu-Yb8* elements in the haploid human genome [23]. Retrotransposition of a few *Alu* sequences has occurred so recently that they are not yet fixed within the human genome [1,18]. The reasons for the differential rate of *Alu* insertion are unknown, which again emphasises how little is understood about retrotransposition [1,24].

The vast majority of *Alu* sequences are believed to be unable to act as templates for further *Alu* expansion within the human genome for two main reasons. Firstly, a large number

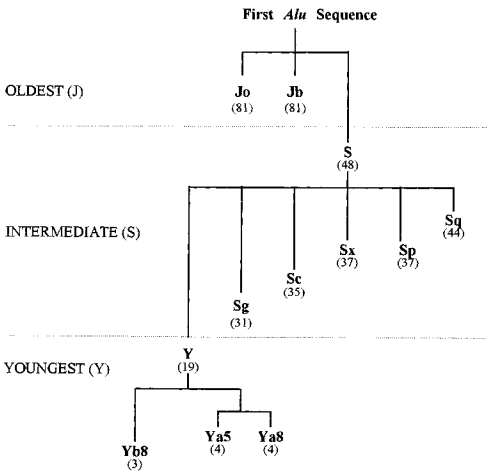

Fig. 2. The proposed evolution of the 12 *Alu* subfamilies. Numbers in parentheses represent approximate times (in millions of years) of insertion of different subfamilies into the human genome. See text for a more complete explanation and related references.

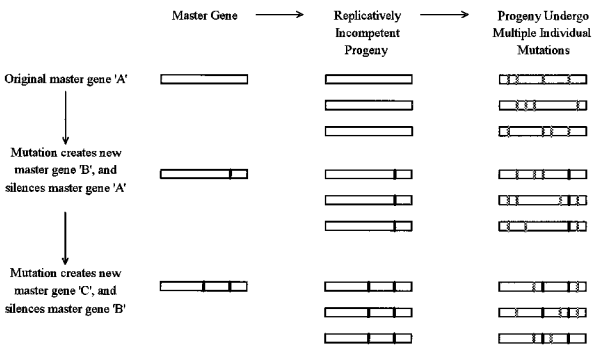

Fig. 3. Proposed evolving master gene model of *Alu* subfamily formation. Adapted from [25]. See text for explanation.

of *Alu* sequences are truncated, particularly at the 5′-end. Secondly, many *Alu* sequences have only 70–80% homology with their respective subfamily consensus sequences [25]. This has lead to formulation of a 'master gene' model of *Alu* formation (Fig. 3), in contrast to the 'random template' model where all *Alu* sequences within the human genome are able to act as templates for *Alu* expansion.

In the master gene model, a highly restricted number of 'master' genes have been retrotransposed many times [25]. A mutation in a master gene generated a new *Alu* subfamily master gene, with subsequent progeny distinguished by this new key mutation. The progeny *Alu* sequences underwent multiple independent mutations, and so diverged from the master gene consensus sequence.

The high rate of independent mutations observed in progeny *Alu* sequences can be attributed to a combination of several factors. Newly retrotransposed *Alu* sequences are CG rich (approximately 65% of their nucleotides), have a high percentage of dCpG doublets compared to other human DNA (9% versus less than 1%) and are extensively methylated [26–28]. Twenty-six dCpG doublets were identified in a recently retrotransposed *Alu* sequence, and it has been estimated that nearly a third of all human dCpG doublets occur within *Alu* sequences [29,30]. Methylated dCpG doublets mutate at a higher rate compared to other nucleotide positions, with 5-methyldeoxycytidine (5-medC) deaminated to generate either dTpG or dCpA [26]. Of relevance to the study of *Alu* sequences was the recent observation that 5-medC deamination is 1250-fold slower in vivo compared to in vitro [31]. Many old *Alu* sequences have lost most of their dCpG motifs [27]. By contrast, *Alu* master genes are assumed not to be extensively methylated as they must be able to conserve their dCpG doublets. This hypothesis is supported by the observation that the BC200 gene, which evolved from FLAM, contains eight dCpG doublets [8,32].

5-medC deamination is an important, but not exclusive cause of *Alu* sequence variation. Random nucleotide mutations probably account for a significant proportion of sequence variability, particularly in older *Alu* sequences. There is much greater drift from the consensus subfamily sequences in old compared to young *Alu* sequences. In addition, there is the potential for the introduction of nucleotide mutations during *Alu* retrotransposition as RNA pol III and reverse transcriptase have higher error rates than DNA polymerases [33]. However, in reality this is probably only a minor source of nucleotide variation.

It is apparent that at several time points during evolution

there was concurrent retrotransposition of more than one *Alu* subfamily. This may indicate either the presence of multiple active master genes, possibly at a single locus, or allelic variation in a single master gene [25,34]. Clearly, only *Alu* sequences retrotransposed in germ cells can be passed from generation to generation. In theory, *Alu* retrotransposition may also occur in somatic cells, but whether this occurs and of what significance it would be remain unclear.

Very rarely, other mechanisms contribute to *Alu* placement within the human genome. For example, comparison of orthologous loci in the low density lipoprotein receptor genes of primates revealed that an old, highly mutated *Alu* repeat present in primates had been replaced by a young *Alu-Yb8* repeat in humans [35]. This represented changes at 16 diagnostic nucleotide positions and included three separate duplication-deletion events. The donor *Alu-Yb8* may have been either single- or double-stranded nucleic acid. Alternatively, a new *Alu-Yb8* repeat may have been inserted into the d(A) tract of the existing old *Alu*, after which a homologous, but unequal crossover event occurred with one of the alleles lost and not passed to the next generation.

4. Functions of *Alu* sequences

The functional role of most *Alu* sequences remains contentious. *Alu* sequences have been considered as either 'junk', 'parasitic' or 'selfish' DNA that served no useful function, yet was not detrimental to the host [36,37]. Conversely, the persistence over a long time period of putative *Alu* master genes in low copy number implied a useful function. Otherwise mutational inactivation and selective clearance from the genome would have occurred [25].

If *Alu* sequences have an important function this implies that non-primates either lack this function, or have compensatory mechanisms. The great variation within copy number in some SINE families among related species such as rat and Chinese hamster suggested that many retrotransposed SINEs do not have a critical role [38]. The genome of the puffer fish (*Fugu rubripes*) is 7.5-fold smaller than the human genome, but many genes are conserved between the two species. The smaller puffer fish genome is the result of shorter intergenic and intronic sequences, and less repetitive DNA. This suggested again that much of the repetitive DNA in humans was redundant [39]. However, within the organised structure of the human genome, it seemed unlikely that all *Alu* sequences were totally functionless and this has now been confirmed.

5. Frequency and distribution of *Alu* sequences

Alu sequences occur at an average of one every 3–6 kbp, but distribution within the human genome is not uniform. *Alu*-rich regions are recognised, yet it remains unknown if these reflect either specific functions or regions where *Alu* sequences exert minimal detrimental effects. For example, just telomeric to the *Bat2* gene in the HLA class III locus, there are 42 *Alu* repeats at a density of 1.9 *Alu* repeats per kbp [40]. Accordingly, there are also regions where *Alu* sequences are sparse.

6. *Alu* sequences are rare in protein-coding exons

Alu repeats are rarely present in protein-coding regions of mature mRNA and controversy has surrounded certain exons that appear to include *Alu* elements [41–44]. *Alu* sequences contain many stop codons in both sense and antisense directions that would result in a truncated protein. One such example is a case of haemophilia B caused by introduction of an *Alu-Ya5* element into a protein coding exon of the factor IX gene [45]. In the few functional mRNA moieties that contain a protein coding *Alu* sequence, the mRNA is often of low abundance compared to other splice variants of the same gene that lack the *Alu* cassette [46,47]. For example, only about 10% of decay accelerating factor mRNA included a protein-coding *Alu-Sc* cassette [48]. The presence of *Alu-J* cassettes in some splice variants of biliary glycoprotein mRNA indicated a long-standing tolerance of protein coding *Alu* elements [49].

7. *Alu* sequences occur frequently in heterogeneous nuclear RNA

Although *Alu* sequences are rare in protein-coding exons, they more commonly occur within the non-coding regions of mature mRNAs. An *Alu* sequence was identified in 5% of 1616 human full-length cDNAs, with 82% and 14% of these located in 3′-UTR and 5′-UTR respectively [44]. In addition, *Alu* sequences frequently occur within introns, and so are present in heterogeneous nuclear RNA. These *Alu* sequences can have a dramatic impact on gene expression. *Alu* sequences, particularly in the antisense direction, contain several regions that differ from either consensus donor or acceptor splice site sequences by either one or two nucleotides [47]. Normal mRNA splicing can be disrupted by point mutation in existing intronic *Alu* elements with activation of cryptic splice sites and subsequent formation of abnormal protein products and clinical disease including Alport syndrome and gyrate atrophy of the choroid and retina [50,51]. Normal mRNA splicing can also be disrupted by retrotransposition of an *Alu* element into an intron, with formation of a truncated, abnormal protein, such as in a reported case of neurofibromatosis [52].

8. *Alu* sequences influence normal gene expression

Alu sequences can influence normal gene expression [53,54]. For example, young subfamily *Alu* sequences include functional retinoic acid response elements that potentially influence expression of several genes including the keratin K18 gene [55]. In another example, an *Alu* sequence within the final intron of the human CD8α gene includes four transcription factor binding sites [56]. An upstream, antisense, monomeric *Alu* sequence in the same intron can bind the downstream complete *Alu* by formation of a cruciform structure, that prevents transcription factor binding, and so silences the CD8α gene enhancer. A significant number of genes include *Alu* sequences in their 5′-region and it is possible that some of these exert an effect on normal transcriptional regulation.

9. *Alu* sequences are involved in chromosome rearrangements

Once retrotransposed, it is thought that few *Alu* sequences are eliminated, although recombinations between distant *Alu* elements can occur [57]. Examples that cause disease such as familial hypercholesterolaemia and angioedema are recognised, with *Alu-Alu* recombinations typically close to the 5′-

end of the left *Alu* monomer [58,59]. Complex events, such as inversion/deletions between *Alu* elements, have also been identified. For example, complex homologous recombination of three intragene *Alu* elements in the GPIIIa gene of the platelet fibronectin receptor has caused Glanzmann's thrombasthenia, a hereditary bleeding disorder [60]. More subtle gene rearrangements also occur, such as polymorphism in the angiotensin-converting enzyme gene, where homozygous absence of a 287 bp intronic *Alu* sequence is associated by meta-analysis with an increased risk of myocardial infarction [61]. Recombination events may involve distant regulatory motifs rather than the downstream introns and exons. For example, deletion of a major, remote regulatory element controlling α-globin expression has caused α-thalassaemia [62]. *Alu* sequences have also been implicated in the generation of pseudogenes [63].

10. *Alu* sequences may have a role in genomic imprinting

Alu sequences are highly methylated in somatic tissues and female germ cells, but methylation in male germ cells is restricted [1,30]. Accordingly, it has been speculated that *Alu* sequences may contribute to the process of genomic imprinting, the process by which maternal and paternal genes are selected for differential expression. Hypomethylation in the male germline may contribute to the preservation of dCpG doublets, for example in the putative *Alu* 'master genes', but the processes of genomic imprinting remain poorly understood.

11. *Alu* sequences as substrates for RNA pol III-mediated transcripts

In addition to *Alu* sequences incorporated in RNA pol II transcripts as discussed earlier, a population of poorly understood RNA pol III transcripts exists which contain *Alu* sequences. Single stranded RNA *Alu* sequences occur in the cytoplasm of transformed and non-transformed human cell lines, and normal human tissues [64–67]. Transport from the nucleus to the cytoplasm of these *Alu* transcripts and 7SL RNA is a competitive, facilitated process dependent upon the *Alu* sequence itself [68].

Two main forms of cytoplasmic RNA pol III *Alu*-containing transcripts have been identified. Full-length *Alu* RNA (fl*Alu* RNA) includes the full-length 3′-polyadenylated *Alu* sequence, and may provide a population of moieties suitable for retrotransposition. Small cytoplasmic *Alu* RNA (sc*Alu* RNA) includes a non-polyadenylated left-hand *Alu* monomer that is probably derived from processed fl*Alu* RNA. BC200 RNA is also very similar to the left-hand *Alu* monomer, but is derived from the BC200 gene. All *Alu* subfamilies are represented in fl*Alu* RNA and sc*Alu* RNA indicating origins from several genomic loci, but young subfamilies predominate. This may reflect the loss of specific control elements, such as a functional RNA pol III promoter as a result of nucleotide mutations, in older *Alu* subfamilies [67,69,70]. Recently, a genomic *Alu-Ya5* sequence with transcriptional activity in vitro was described, with transcription partially dependent upon the flanking sequences [70].

Despite the high numbers of *Alu* repeats in the human genome each cell typically includes less than 1000 copies each of fl*Alu* RNA and sc*Alu* RNA [69]. However, the abundance of cytoplasmic *Alu* transcripts in vitro can be modified by either viral infection, alterations in nucleotide methylation, heat shock or chemical manipulation. Cellular infection with a number of viruses caused transient accumulation of high levels of cytoplasmic fl*Alu* RNA transcripts [71–73]. Viral gene expression was necessary, and increased RNA pol III transcription is believed to reflect roles for certain viral proteins as associated transcription factors. Speculation has occurred about an anti-viral biological role for *Alu* transcripts. The *Alu* sequence is derived from, and retains a similar secondary structure to, the functional domain of 7SL RNA that is responsible for the protein elongation-arresting function of the SRP [14]. 7SL becomes a functional part of the SRP by interacting with specific proteins, and *Alu* sequences are also able to bind proteins homologous to components of the SRP [74–77]. Accordingly, *Alu* transcripts complexed with specific proteins may have a 7SL-like function. For example, in viral-infected cells it has been postulated that *Alu* transcripts complexed with protein could arrest viral and host protein synthesis, with prevention of viral replication and death of infected host cells in a mechanism reminiscent of that induced by interferon [27]. HeLa cell proliferation can be inhibited by either 7SL or *Alu* gene sequences in expression vectors [78].

Regulation of these *Alu*-containing RNA pol III transcripts is likely to be influenced by several different mechanisms that are currently poorly understood. For example, a recent report described a level of control that involved wild-type p53 repression of *Alu* template activity in vitro and in vivo [79]. Further insights into the function of *Alu*-containing RNA pol III transcripts may eventually be gained by greater understanding of the function of BC200 RNA [8,16].

12. Conclusions

Researchers investigating the human genome inevitably encounter *Alu* sequences. The opinion that the large number of *Alu* sequences within our genome represent the consequences of 'parasitic' expansion of retrotransposons with no benefit to the human host has been repeatedly challenged in recent years. However, many aspects of *Alu* sequence function remain contentious. A large number of *Alu* repeats may be 'junk' DNA that confers neither benefit nor harm to the host. However, specific adaptive and maladaptive functions have been identified for an increasing number of individual *Alu* sequences. Some of these functions are critically dependent on the nucleotide sequences of specific *Alu* cassettes, rather than reflecting a defined, fundamental role for all *Alu* sequences. It seems probable that further functions for *Alu* sequences will be identified in the future.

Acknowledgements: A.J.M. thanks the Medical Research Council and the Foulkes Foundation for specific support. Research in the authors' laboratory is also funded by the Wellcome Trust, Yorkshire Cancer Research Campaign, Northern Regional Health Authority and West Riding Medical Research Trust.

References

[1] Schmid, C.W. (1996) Prog. Nucleic Acid Res. Mol. Biol. 53, 283–319.
[2] Houck, C.M., Rinehart, F.P. and Schmid, C.W. (1979) J. Mol. Biol. 132, 289–306.

[3] Fuhrman, S.A., Deininger, P.L., LaPorte, P., Friedmann, T. and Geiduschek, E.P. (1981) Nucleic Acids Res. 9, 6439–6456.
[4] Willis, I.M. (1993) Eur. J. Biochem. 212, 1–11.
[5] Hutchinson, G.B., Andrew, S.E., McDonald, H., Goldberg, Y.P., Graham, R., Rommens, J.M. and Hayden, M.R. (1993) Nucleic Acids Res. 21, 3379–3383.
[6] Jurka, J. and Klonowski, P. (1996) J. Mol. Evol. 43, 685–689.
[7] Quentin, Y. (1994) Nucleic Acids Res. 22, 2222–2227.
[8] Quentin, Y. (1994) Genetica 93, 203–215.
[9] Weiner, A.M., Deininger, P.L. and Efstratiadis, A. (1986) Annu. Rev. Biochem. 55, 631–661.
[10] Deininger, P.L. (1989) in: Mobile DNA (Berg, D.E. and Howe, M.M., Eds.), 1st edn., pp. 619–636, American Society of Microbiology, Washington, DC.
[11] Bucala, R., Lee, A.T., Rourke, L. and Cerami, A. (1993) Proc. Natl. Acad. Sci. USA 90, 2666–2670.
[12] Maraia, R.J., Kenan, D.J. and Keene, J.D. (1994) Mol. Cell. Biol. 14, 2147–2158.
[13] Boeke, J.D. (1997) Nature Genet. 16, 6–7.
[14] Labuda, D. and Zietkiewicz, E. (1994) J. Mol. Evol. 39, 506–518.
[15] Quentin, Y. (1992) Nucleic Acids Res. 20, 3397–3401.
[16] Martignetti, J.A. and Brosius, J. (1993) Proc. Natl. Acad. Sci. USA 90, 11563–11567.
[17] Quentin, Y. (1992) Nucleic Acids Res. 20, 487–493.
[18] Batzer, M.A., Deininger, P.L., Hellmann-Blumberg, U., Jurka, J., Labuda, D., Rubin, C.M., Schmid, C.W., Zietkiewicz, E. and Zuckerkandl, E. (1996) J. Mol. Evol. 42, 3–6.
[19] Bailey, A.D. and Shen, C.-K.J. (1993) Proc. Natl. Acad. Sci. USA 90, 7205–7209.
[20] Britten, R.J. (1994) Proc. Natl. Acad. Sci. USA 91, 6148–6150.
[21] Kapitonov, V. and Jurka, J. (1996) J. Mol. Evol. 42, 59–65.
[22] Shaikh, T.H. and Deininger, P.L. (1996) J. Mol. Evol. 42, 15–21.
[23] Zietkiewicz, E., Richer, C., Makalowski, W., Jurka, J. and Labuda, D. (1994) Nucleic Acids Res. 22, 5608–5612.
[24] Schmid, C.W. and Maraia, R. (1992) Curr. Opin. Genet. Dev. 2, 874–882.
[25] Deininger, P.L., Batzer, M.A., Hutchison, C.A. and Edgell, M.H. (1992) Trends Genet. 8, 307–311.
[26] Bird, A.P. (1980) Nucleic Acids Res. 8, 1499–1504.
[27] Jurka, J. and Milosavljevic, A. (1991) J. Mol. Evol. 32, 105–121.
[28] Schmid, C.W. (1991) Nucleic Acids Res. 19, 5613–5617.
[29] Stoppa-Lyonnet, D., Carter, P.E., Meo, T. and Tosi, M. (1990) Proc. Natl. Acad. Sci. USA 87, 1551–1555.
[30] Hellmann-Blumberg, U., Hintz, M.F., Gatewood, J.M. and Schmid, C.W. (1993) Mol. Cell. Biol. 13, 4523–4530.
[31] Yang, A.S., Gonzalgo, M.L., Zingg, J.-M., Millar, R.P., Buckley, J.D. and Jones, P.A. (1996) J. Mol. Biol. 258, 240–250.
[32] Labuda, D. and Striker, G. (1989) Nucleic Acids Res. 17, 2477–2491.
[33] Varmus, H. and Brown, P. (1989) in: Mobile DNA (Berg, D.E. and Howe, M.H., Eds.), 1st edn,, pp. 53–108, American Society of Microbiology, Washington, DC.
[34] Shen, M.R., Batzer, M.A. and Deininger, P.L. (1991) J. Mol. Evol. 33, 311–320.
[35] Kass, D.H., Batzer, M.A. and Deininger, P.L. (1995) Mol. Cell. Biol. 15, 19–25.
[36] Doolittle, W.F. and Sapienza, C. (1980) Nature 284, 601–603.
[37] Orgel, L.E. and Crick, F.H.C. (1980) Nature 284, 604–607.
[38] Sapienza, C. and St-Jacques, B. (1986) Nature 319, 418–420.
[39] Baxendale, S., Abdulla, S., Elgar, G., Buck, D., Berks, M., Micklem, G., Durbin, R., Bates, G., Brenner, S., Beck, S. and Lehrach, H. (1995) Nature Genet. 10, 67–75.
[40] Iris, F.J.M., Bougueleret, L., Prieur, S., Caterina, D., Primas, G., Perrot, V., Jurka, J., Rodriguez-Tome, P., Claverie, J.M., Dausset, J. and Cohen, D. (1993) Nature Genet. 3, 137–145.
[41] Yeo, J.-P., Alderuccio, F. and Toh, B.-H. (1994) Nature 367, 288–291.
[42] Tugendreich, S., Feng, Q., Kroll, J., Sears, D.D., Boeke, J.D. and Hieter, P. (1994) Nature 370, 106.
[43] Margalit, H., Nadir, E. and Ben-Sasson, S.A. (1994) Cell 78, 173–174.
[44] Yulug, I.G., Yulug, A. and Fisher, E.M.C. (1995) Genomics 27, 544–548.
[45] Vidaud, D., Vidaud, M., Bahnak, B.R., Siguret, V., Sanchez, S., Laurian, Y., Meyer, D., Goosens, M. and Lavergne, J.M. (1993) Eur. J. Hum. Genet. 1, 30–36.
[46] Claverie, J.-M. (1992) Genomics 12, 838–841.
[47] Makalowski, W., Mitchell, G.A. and Labuda, D. (1994) Trends Genet. 10, 188–193.
[48] Caras, I.W., Davitz, M.A., Rhee, L. and Nussenzweig, V. (1987) Nature 325, 545–549.
[49] Barnett, T.R., Drake, L. and Pickle, W. (1993) Mol. Cell. Biol. 13, 1273–1282.
[50] Mitchell, G.A., Labuda, D., Fontaine, G., Saudubray, J.M., Bonnefont, J.P., Lyonnet, S., Brody, L.C., Steel, G., Obie, C. and Valle, D. (1991) Proc. Natl. Acad. Sci. USA 88, 815–819.
[51] Knebelmann, B., Forestier, L., Drouot, L., Quinones, S., Chuet, C., Benessy, F., Saus, J. and Antignac, C. (1995) Hum. Mol. Genet 4, 675–679.
[52] Wallace, M.R., Andersen, L.B., Saulino, A.M., Gregory, P.E., Glover, T.W. and Collins, F.S. (1991) Nature 353, 864–866.
[53] Britten, R.J. (1996) Proc. Natl. Acad. Sci. USA 93, 9374–9377.
[54] Almenoff, J.S., Jurka, J. and Schoolnik, G.K. (1994) J. Biol. Chem. 269, 16610–16617.
[55] Vansant, G. and Reynolds, W.F. (1995) Proc. Natl. Acad. Sci. USA 92, 8229–8233.
[56] Hanke, J.H., Hambor, J.E. and Kavathas, P. (1995) J. Mol. Biol. 246, 63–73.
[57] Renwick, P.J., Birley, A.J. and Hulten, M.A. (1997) Gene 184, 155–162.
[58] Stoppa-Lyonnet, D., Duponchel, C., Meo, T., Laurent, J., Carter, P.E., Arala-Chaves, M., Cohen, J.H., Dewald, G., Goetz, J., Hauptmann, G., Lagre, G., Lesavre, P., Lopez-Trascasa, M., Misiano, G., Moraine, C., Sobel, A., Späth, P.J. and Tosi, M. (1991) Am. J. Hum. Genet. 49, 1055–1062.
[59] Chae, J.J., Park, Y.B., Kim, S.H., Hong, S.S., Song, G.J., Han, K.H., Namkoong, Y., Kim, H.S. and Lee, C.C. (1997) Hum. Genet. 99, 155–163.
[60] Li, L. and Bray, P.F. (1993) Am. J. Hum. Genet. 53, 140–149.
[61] Samani, N.J., Thompson, J.R., O'Toole, L., Channer, K. and Woods, K.L. (1996) Circulation 94, 708–712.
[62] Flint, J., Rochette, J., Craddock, C.F., Dode, C., Vignes, B., Horsley, S.W., Kearney, L., Buckle, V.J., Ayyub, H. and Higgs, D.R. (1996) Hum. Mol. Genet. 5, 1163–1169.
[63] Marchese, A., Beischlag, T.V., Nguyen, T., Niznik, H.B., Weinshank, R.L., George, S.R. and O'Dowd, B.F. (1995) Gene 154, 153–158.
[64] Matera, A.G., Hellmann, U. and Schmid, C.W. (1990) Mol. Cell. Biol. 10, 5424–5432.
[65] Sinnett, D., Richer, C., Deragon, J.-M. and Labuda, D. (1992) J. Mol. Biol. 226, 689–706.
[66] Maraia, R.J., Driscoll, C.T., Bilyeu, T., Hsu, K. and Darlington, G.J. (1993) Mol. Cell. Biol. 13, 4233–4241.
[67] Liu, W.-M., Maraia, R.J., Rubin, C.M. and Schmid, C.W. (1994) Nucleic Acids Res. 22, 1087–1095.
[68] He, X.-P., Bataillé, N. and Fried, H.M. (1994) J. Cell. Sci. 107, 903–912.
[69] Liu, W.-M. and Schmid, C.W. (1993) Nucleic Acids Res. 21, 1351–1359.
[70] Shaikh, T.H., Roy, A.M., Kim, J., Batzer, M.A. and Deininger, P.L. (1997) J. Mol. Biol. 271, 222–234.
[71] Jang, K.L., Collins, M.K.L. and Latchman, D.S. (1992) J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 5, 1142–1147.
[72] Panning, B. and Smiley, J.R. (1993) Mol. Cell. Biol. 13, 3231–3244.
[73] Panning, B. and Smiley, J.R. (1994) Virology 202, 408–417.
[74] Chang, D.-Y., Hsu, K. and Maraia, R.J. (1996) Nucleic Acids Res. 24, 4165–4170.
[75] Chiang, Y. and Vishwanatha, J.K. (1996) Mol. Cell. Biochem. 155, 131–138.
[76] Bovia, F., Wolff, N., Ryser, S. and Strub, K. (1997) Nucleic Acids Res. 25, 318–326.
[77] Sarrowa, J., Chang, D.-Y. and Maraia, R.J. (1997) Mol. Cell. Biol. 17, 1144–1151.
[78] Sakamoto, K., Fordis, C.M., Corsico, C.D., Howard, T.H. and Howard, B.H. (1991) J. Biol. Chem. 266, 3031–3038.
[79] Chesnokov, I., Chu, W.M., Botchan, M.R. and Schmid, C.W. (1996) Mol. Cell. Biol. 16, 7084–7088.