J Mol Evol (1996) 42:3–6

JOURNAL OF MOLECULAR EVOLUTION

Standardized Nomenclature for Alu Repeats

Mark A. Batzer,[1]* Prescott L. Deininger,[2] Utha Hellmann-Blumberg,[3] Jerzy Jurka,[4] Damian Labuda,[5] Carol M. Rubin,[6] Carl W. Schmid,[3,6] Ewa Ziętkiewicz,[5] Emile Zuckerkandl[7]

[1] Human Genome Center, L-452, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94551, USA
[2] Department of Biochemistry and Molecular Biology, Center for Human and Molecular Genetics, Louisiana State University Medical Center, 1901 Perdido St., New Orleans, LA 70112, USA
[3] Department of Chemistry, University of California at Davis, Davis, CA 95616, USA
[4] Linus Pauling Institute of Science and Medicine, 440 Page Mill Road, Palo Alto, CA 94306, USA
[5] Centre de Recherche, Hôpital Ste-Justine, Departement de Pediatrie, Universite de Montréal, Montréal, Quëbec, Canada H3T 1C5
[6] Section of Molecular and Cell Biology, University of California at Davis, Davis, CA 95616, USA
[7] Institute of Molecular Medical Sciences, 460 Page Mill Road, Palo Alto, CA 94306, USA

Received: 19 January 1995 / Accepted: 23 February 1995

Short interspersed elements (SINEs) may be found in the genomes of a wide variety of mammals (Deininger and Batzer 1993). The Alu family of SINEs is one of the most successful mobile genetic elements, having arisen to a copy number in excess of 500,000 within the human genome in approximately 65 million years of primate evolution. (For reviews see Deininger 1989; Okada 1991; Schmid and Maraia 1992; Deininger and Batzer 1993.) Alu sequences are thought to be ancestrally derived from the 7SL RNA gene (Ullu et al. 1982) and to mobilize through an RNA polymerase III–derived transcript in a process termed retroposition (Rogers 1983).

Alu sequences within the human genome can be divided into subfamilies of related elements based upon diagnostic mutations shared by subfamily members. Several groups have independently identified a series of overlapping subfamilies of Alu repeats (Slagel et al. 1987; Willard et al. 1987; Britten et al. 1988; Jurka and Smith 1988; Quentin 1988; Deininger and Slagel 1988; Shen et al. 1991; Jurka and Milosavljevic 1991; Batzer et al. 1990; Matera et al. 1990a,b; Batzer and Deininger 1991; Jurka 1993; Hutchinson et al. 1993) which appear to be of different genetic ages (Slagel et al. 1987; Willard et al. 1987; Britten et al. 1988; Jurka and Smith 1988; Quentin 1988; Deininger and Slagel 1988; Labuda and Striker, 1989; Shen et al. 1991; Jurka and Milosavljevic 1991; Batzer et al. 1990; Matera et al. 1990a,b; Batzer and Deininger 1991; Jurka 1993; Hutchinson et al. 1993; Britten 1994). These observations have led to the suggestion that the vast majority of Alu amplifications were derived from a small subset of active Alu ''master'' genes (Deininger et al. 1992). However, it is clear that several Alu subfamilies are currently undergoing amplification from multiple ''master'' genes within the human genome (Matera et al. 1990b; Leeflang et al. 1992; Jurka 1993; Hutchinson et al. 1993; Hammer 1994; Batzer et al. 1995; Deininger and Batzer, 1995).

The younger subfamilies of Alu sequences contain individual members that are restricted to the human genome, some of which are polymorphic (Deininger and Slagel 1988; Batzer et al. 1990; Matera et al. 1990b; Stoppa-Lyonnet et al. 1990; Batzer and Deininger 1991; Batzer et al. 1991, 1994, 1995, 1996; Muratani et al. 1991; Perna et al. 1992; Hutchinson et al. 1993; Jurka 1993; Blonden et al. 1994; Hammer 1994; Kass et al. 1994). Some of the young subfamily members have arisen as the result of de novo insertions in the NF-1 gene (Wallace et al. 1991) or factor IX gene (Vidaud et al. 1993) within the human genome, clearly demonstrating

* *Present address:* Department of Pathology, Stanley S. Scott Cancer Center, Louisiana State University Medical Center, 1901 Perdido St., New Orleans, LA 70112.
Correspondence to: M.A. Batzer

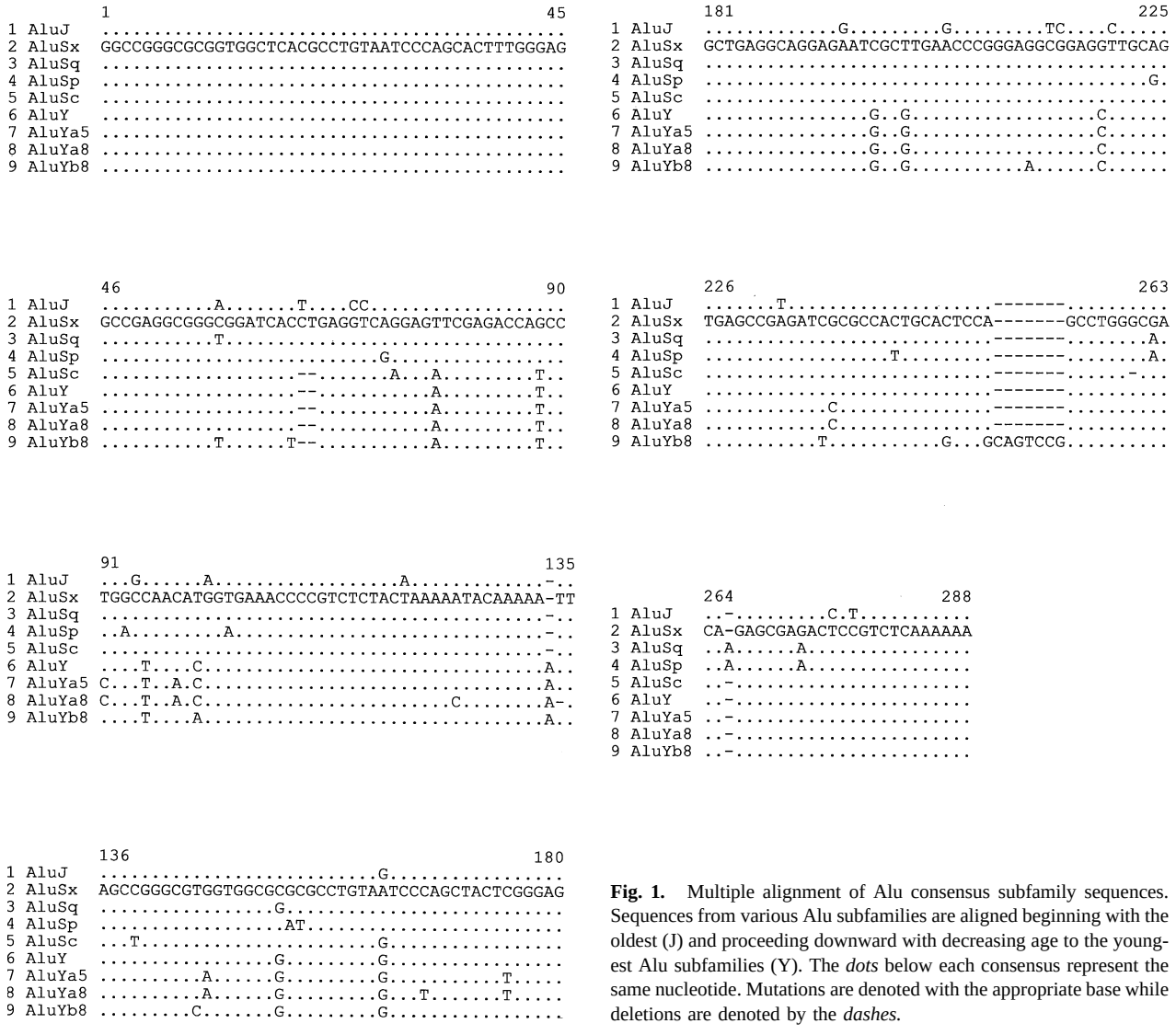

Fig. 1. Multiple alignment of Alu consensus subfamily sequences. Sequences from various Alu subfamilies are aligned beginning with the oldest (J) and proceeding downward with decreasing age to the youngest Alu subfamilies (Y). The *dots* below each consensus represent the same nucleotide. Mutations are denoted with the appropriate base while deletions are denoted by the *dashes.*

that these elements are still actively inserted within humans. The identification of young Alu repeats provides new information about the biological properties of these elements as well as genomic fossils for the study of human genetic diversity. Therefore, the recent or young Alu insertions represent a very important group of Alu repeats dispersed throughout the human genome.

The nomenclature describing Alu repeats is not standardized. As an ever-increasing proportion of the human genome is sequenced it has become imperative to derive an accepted nomenclature for classifying newly sequenced Alu family members. We have attempted to combine some of the preexisting nomenclature with newer names to provide a basis for naming Alu sequences. In the universal nomenclature a logical (alphabetical) progression from the oldest (J) to intermediate (S) and young (Y) Alu sequences using capital letters is used to denote major subfamily branches (Fig. 1). The primary nomenclature to be used for the two older classes of repeats (J and S) has been described previously (Jurka and Smith 1988; Jurka and Milosavljevic 1991), along with a series of subbranches of the S family (x, p, q, c). The intermediate and older subfamilies all have a significant amount of heterogeneity, and there are also many examples of intermediates between these various subfamilies. Thus, this selection is not meant to be an exhaustive selection of all possible subfamilies, but simply a reasonable working nomenclature for those older subfamilies.

With the younger subfamilies, there is less general heterogeneity but still the possibility of finding intermediates between the previously identified subfamilies. The current scheme utilizes lineages starting from the Y subfamily consensus sequence. The Y subfamily is considered a ‘‘gold standard’’ since it has been previously identified as a subfamily by a number of different laboratories (outlined in the synonyms below). All Alu repeats presently known to retropose differ from the Y

subfamily consensus sequence by only a few additional diagnostic mutations, suggesting that the younger subfamilies of Alu repeats were ancestrally derived from the Y subfamily. Therefore, young subfamilies are defined as lineages that descended from this gold standard. They are defined by a lowercase letter chosen in alphabetical order (a, b, c . . .) based solely on the order of publication of the different lineages. In addition to the lineage, a number is assigned to each subfamily based on the number of diagnostic changes relative to the Y consensus. Thus, the HS/PV subfamily becomes Ya5, because it was the first Y subfamily lineage defined and has five diagnostic changes. The Ya8 subfamily is in the same lineage but has three additional changes. Minor variants of the major designations which are identified in nonhuman primates will be noted by the addition of an abbreviated name in italics that denotes the genus and species of the primate (e.g., a Ya5 variant from the common chimpanzee, *Pan troglodytes* becomes Ya5*Ptr* etc.).

This nomenclature was designed and derived during the conference ''SINEs, LINEs and Retrotransposable Elements: Functional Implications'' held in the summer of 1994. It was agreed upon by all of the authors of this manuscript and will be strictly adhered to by all the authors. Shown below are samples of the new nomenclature and their older synonyms.

- AluY → CS (Shen et al. 1991), Sub (Slagel et al. 1987), Conserved (Willard et al. 1987), Precise (Britten et al. 1989), Class IV (Britten et al. 1988), Type A (Quentin 1988), Sb (Jurka and Smith 1988; Jurka and Milosavljevic 1991)
- AluYa5 → 'new' (Deininger and Slagel 1988), HS (Batzer et al. 1990; Batzer and Deininger 1991), PV (Matera et al. 1990a, b), Sb1 (Jurka 1993)
- AluYa8 → HS-2 (Batzer et al. 1990; Batzer and Deininger 1991; Batzer et al. 1995), PV minor (Matera et al. 1990b)
- AluYb8 → Sb2 (Jurka 1993; Hutchinson et al. 1993; Zietkiewicz et al. 1994; Batzer et al. 1995)

Acknowledgments. The nomenclature described here was developed as part of the conference ''SINEs, LINEs and Retrotransposable Elements: Functional Implications'' sponsored by the International Society of Molecular Evolution and funded by the Alfred P. Sloan Foundation and the United States Department of Energy. This research was supported by grants from the U.S. Department of Energy (LDRD 94-LW-103) to M.A.B. and (DE-FG03-91ER61152) to J.J.; from Cancer Research Society, Inc. to D.L.; Medical Research Council of Canada (MA-12175) to D.L.; and from the National Institutes of Health (RO1 HG 00770) to P.L.D. and (GM 21346) to C.W.S. Work at Lawrence Livermore National Laboratory was performed under the auspices of U.S. Department of Energy contract No. W-7405-ENG-48.

References

Batzer MA, Deininger PL (1991) A human-specific subfamily of Alu sequences. Genomics 9:481–487

Batzer MA, Kilroy GE, Richard PE, Shaikh TH, Desselle TD, Hoppens CL, Deininger PL (1990) Structure and variability of recently inserted Alu family members. Nucleic Acids Res 18:6793–6798

Batzer MA, Gudi VA, Mena JC, Foltz DW, Herrera RJ, Deininger PL (1991) Amplification dynamics of Human-Specific (HS) Alu family members. Nucleic Acids Res 19:3619–3623

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, Deininger PL (1994) African origin of human-specific polymorphic Alu insertions. Proc Natl Acad Sci USA 91:12288–12292

Batzer MA, Arcot SS, Phinney JW, Alegria-Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, Hochmeister M, Ioannou PA, Herrera RJ, Boudreau DA, Scheer WD, Keats BJB, Deininger PL, Stoneking M (1996) Genetic variation of recent Alu insertions in human populations. J Mol Evol 42:22–29

Batzer MA, Rubin CM, Hellman-Blumberg U, Alegria-Hartman M, Leeflang EP, Stern JD, Bazan HA, Shaikh TH, Deininger PL, Schmid CW (1995) Dispersion and insertion polymorphism in two small subfamilies of recently amplified human Alu repeats. J Mol Biol 247:418–427

Blonden LAJ, Terwindt GM, Den Dunnen JT, Van Ommen G-JB (1994) A polymorphic STS in intron 44 of the dystrophin gene. Hum Genet 93:479–480

Britten RJ (1994) Evidence that most human Alu sequences were inserted in a process that ceased about 30 million years ago. Proc Natl Acad Sci USA 91:6148–6150

Britten RJ, Baron WF, Stout DB, Davidson EH (1988) Sources and evolution of human Alu repeated sequences. Proc Natl Acad Sci USA 85:4770–4774

Britten RJ, Stout DB, Davidson EH (1989) The current source of human Alu retroposons is a conserved gene shared with Old World monkey. Proc Natl Acad Sci USA 86:3718–3722

Deininger PL (1989) SINEs: short interspersed repeated DNA elements in higher eucaryotes. In: Berg DE, Howe MM (eds) Mobile DNA. American Society for Microbiology, Washington, DC, pp 619–636

Deininger PL, Batzer MA (1993) Evolution of retroposons. Evol Biol 27:157–196

Deininger PL, Batzer MA (1995) SINE master genes and population biology. In: Maraia RJ (ed) The impact of short interspersed elements (SINEs) on the host genome. RG Landes, Georgetown, TX, pp 43–60

Deininger PL, Slagel VK (1988) Recently amplified Alu family members share a common parental Alu sequence. Mol Cell Biol 8:4566–4569

Deininger PL, Batzer MA, Hutchison III CA, Edgell MH (1992) Master genes in mammalian repetitive DNA amplification. Trends Genet 8:307–311

Hammer MF (1994) A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. Mol Biol Evol 11:749–761

Hutchinson GB, Andrew SE, McDonald H, Goldberg YP, Graharn R, Rommens JR, Hayden MR (1993) An Alu element retroposition in two families with Huntington disease defines a new active Alu subfamily. Nucleic Acids Res 21:3379–3383

Jurka J (1993) A new subfamily of recently retroposed Alu repeats. Nucleic Acids Res 21:2252

Jurka J, Milosavljevic A (1991) Reconstruction and analysis of human Alu genes. J Mol Evol 32:105–121

Jurka J, Smith T (1988) A fundamental division in the Alu family of repeated sequences. Proc Natl Acad Sci USA 85:4775–4778

Kass DH, Aleman C, Batzer MA, Deininger PL (1994) Identification of a human specific Alu insertion in the Factor XIIIB gene. Genetica 94:1–8

Labuda D, Striker G (1989) Sequence conservation in Alu evolution. Nucleic Acids Res 17:2477–2491

Leeflang EP, Liu W–M, Hashimoto C, Choudary PV, Schmid CW (1992) Phylogenetic evidence for multiple Alu source genes. J Mol Evol 35:7–16

Matera AG, Hellmann U, Schmid CW (1990a) A transpositionally and transcriptionally competent Alu subfamily. Mol Cell Biol 10:5424–5432

Matera AG, Hellmann U, Hintz MF, Schmid CW (1990b) Recently transposed Alu repeats result from multiple source genes. Nucleic Acids Res 18:6019–6023

Muratani K, Hada T, Yamamoto Y, Kaneko T, Shigeto Y, Ohue T, Furuyama J, Higashino K (1991) Inactivation of the cholinesterase gene by Alu insertion: possible mechanism for human gene transposition. Proc Natl Acad Sci USA 88:11315–11319

Okada N (1991) SINEs. Curr Opin Genet Dev 1:498–504.

Perna NT, Batzer MA, Deininger PL, Stoneking M (1992) Alu insertion polymorphism: A new type of marker for human population studies. Hum Biol 64:641–648

Quentin Y (1988) The Alu family developed through successive waves of fixation closely connected with primate lineage history. J Mol Evol 27:194–202

Rogers J (1983) Retroposons defined. Nature 301:460

Schmid CW, Maraia R (1992) Transcriptional regulation and transpositional selection of active SINE sequences. Curr Opin Genet Dev 2:874–882.

Shen M-R, Batzer MA, Deininger PL (1991) Evolution of the master Alu gene(s). J Mol Evol 33:311–320

Slagel V, Flemington E, Traina-Dorge V, Bradshaw H, Deininger PL (1987) Clustering and sub-family relationships of the Alu family in the human genome. Mol Biol Evol 14:19–29

Stoppa-Lynnet D, Carter PE, Meo T, Tosi M (1990) Clusters of intragenic Alu repeats predispose the human C1 inhibitor locus to deleterious rearrangements. Proc Natl Acad Sci USA 87:1551–1555

Ullu E, Murphy S, Melli M (1982) Human 7S RNA consists of a 140 nucleotide middle repetitive sequence inserted in an Alu sequence. Cell 29:195–202

Vidaud D, Vidaud M, Bahnak BR, Siguret V, Sanchez SG, Laurin Y, Meyer D, Goossens M, Lavergne JM (1993) Hemophilia B due to a *de novo* insertion of a Human-Specific Alu subfamily member within the coding region of the factor IX gene. Eur J Hum Genet 1:30–36

Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, Collins FS (1991) A *de novo* Alu insertion results in neurofibromatosis type 1. Nature 353:864–866

Willard C, Nguyen HT, Schmid CW (1987) Existence of at least three distinct Alu subfamilies. J Mol Evol 26:180–186

Zietkiewicz E, Richer C, Makalowski W, Jurka J, Labuda D (1994) A young Alu subfamily amplified independently in human and African great apes lineages. Nucleic Acids Res 22:5608–5612