7-1-2007

In search of polymorphic Alu insertions with restricted geographic distributions

Richard Cordaux
Center for BioModular Multi-Scale Systems

Deepa Srikanta
Center for BioModular Multi-Scale Systems

Jungnam Lee
Center for BioModular Multi-Scale Systems

Mark Stoneking
Max Planck Institute for Evolutionary Anthropology

Mark A. Batzer
Center for BioModular Multi-Scale Systems

Recommended Citation

Cordaux, R., Srikanta, D., Lee, J., Stoneking, M., & Batzer, M. (2007). In search of polymorphic Alu insertions with restricted geographic distributions. *Genomics, 90* (1), 154-158. https://doi.org/10.1016/j.ygeno.2007.03.010

Available online at www.sciencedirect.com
ScienceDirect
Genomics 90 (2007) 154–158
GENOMICS
www.elsevier.com/locate/ygeno

Short Communication

In search of polymorphic *Alu* insertions with restricted geographic distributions

Richard Cordaux [a,b], Deepa Srikanta [a], Jungnam Lee [a], Mark Stoneking [c], Mark A. Batzer [a,*]

[a] *Department of Biological Sciences, Biological Computation and Visualization Center, Center for BioModular Multi-Scale Systems, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803, USA*
[b] *Laboratoire de Génétique et Biologie des Populations de Crustacés, CNRS UMR 6556, Université de Poitiers, 40 Avenue du Recteur Pineau, 86022 Poitiers, France*
[c] *Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany*

Received 9 February 2007; accepted 17 March 2007
Available online 11 May 2007

Abstract

Alu elements are transposable elements that have reached over one million copies in the human genome. Some *Alu* elements inserted in the genome so recently that they are still polymorphic for insertion presence or absence in human populations. Recently, there has been an increasing interest in using *Alu* variation for studies of human population genetic structure and inference of individual geographic origin. Currently, this requires a high number of *Alu* loci. Here, we used a linker-mediated polymerase chain reaction method to preferentially identify low-frequency *Alu* elements in various human DNA samples with different geographic origins. The candidate *Alu* loci were subsequently genotyped in 18 worldwide human populations (~370 individuals), resulting in the identification of two new *Alu* insertions restricted to populations of African ancestry. Our results suggest that it may ultimately become possible to correctly infer the geographic affiliation of unknown samples with high levels of confidence without having to genotype as many as 100 *Alu* loci. This is desirable if *Alu* insertion polymorphisms are to be used for human evolution studies or forensic applications.

Keywords: *Alu* insertions; Humans; Polymorphism; Genetic structure; Geographic inference

Alu elements are ~300-bp-long transposable elements that have expanded in primate genomes within the last ~65 million years (My) [1]. *Alu* elements mobilize (i.e., produce new copies) via a "copy and paste" mechanism, in which the RNA transcript of an active element is reverse transcribed as cDNA and the duplicate element is inserted at a new genomic location [1]. Although only a subset of all *Alu* elements are capable of producing new copies [2,3], they continue to expand in the human genome at a substantial rate [4]. As a result, *Alu* elements have reached over one million copies in the human genome, making them the most successful mobile elements in the human genome [1,5]. Concomitantly, *Alu* elements have had a considerable impact on their host genome, e.g., by inducing genetic disease [6] or promoting genomic plasticity [7–9]. Therefore, *Alu* elements represent an important source of human genomic variation [1].

Some *Alu* elements have inserted in the genome so recently that they are still polymorphic for insertion presence or absence at the individual or population levels [1,10]. Because recently integrated *Alu* elements follow a neutral model of evolution [11], they represent an important source of genetic markers for human population studies [1]. Hence, *Alu* elements have proven useful for addressing questions related to human evolution [12–15]. More recently, there has been an increasing interest in using *Alu* variation for studies of human population genetic structure and inference of individual geographic origin [16,17]. For example, by genotyping 100 polymorphic *Alu* loci, Ray et al. [17] were able to correctly infer the geographic affiliation of 18 unknown human individuals with high levels of confidence. However, as noted by the authors, the discovery and characterization of novel *Alu* insertions with restricted geographic distributions may allow reducing the number of elements required without loss of

* Corresponding author. Fax: +1 (225) 578 7113.
E-mail address: mbatzer@lsu.edu (M.A. Batzer).

0888-7543/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2007.03.010

confidence in the results, which could be desirable for applications such as forensics [17].

Previously, it has been technically challenging to identify recently integrated *Alu* elements due to the difficulty of detecting one new insertion among one million preexisting elements in the genome. More recently, computationally based approaches have facilitated the identification of a number of elements differentially inserted among individuals or populations [10,18–21]. Obviously, such approaches are limited by the availability of genomic sequence data. This may result for example in the preferential recovery of high-frequency elements since the elements have to be present in the sequence to be identified. Moreover, the geographic origin or ethnicity of the samples used to generate genomic data may sometimes be unknown or vague, thus rendering the identification and characterization of *Alu* insertions with restricted geographic distributions more difficult. To circumvent these potential disadvantages, we used a linker-mediated PCR-like method previously designed to recover newly inserted *Alu* elements [22], modified to target low-frequency *Alu* elements in various samples with different geographic origins. The candidate *Alu* insertion loci were subsequently genotyped in a panel of worldwide human populations, resulting in the identification of several *Alu* insertions with putatively restricted geographic distributions. These novel genetic markers may prove useful for human evolution studies or forensics applications.

Results and discussion

We searched for new *Alu* insertion loci in two separate individuals and five pools of three individuals with different geographic origins (Table 1). For each set of experiments, we sequenced from ~200 to ~500 different clones. We identified a total of nine candidate *Alu* loci that were amenable to PCR (Table 1). However, one locus (RC3) was recovered twice independently, in the African sample L945 and the African pooled sample. Therefore, there were eight potential new *Alu* elements for further analysis.

Table 1
Summary of display cloning and sequencing results

Continental ancestry	Sample ID (origin)	No. of clones sequenced	No. of *Alu* loci lacking from database	No. of *Alu* loci amenable to PCR
African sample	L945 (African American)	~500	6	4
European sample	10237 (Germany)	~200	1	0
Asian pool 1	10540 (Melanesia) 11377 (Cambodia) 17018 (China)	~500	5	0
Asian pool 2	10539 (Melanesia) 10541 (Melanesia) 10543 (Melanesia)	~500	5	0
European pool 1	13911 (Russia) 13626 (Russia) 10408 (Germany)	~500	4	1
European pool 2	15885 (Basque) 13911 (Russia) 1052 (Germany)	~400	3	2
African pool	19038 (Kenya) 19046 (Kenya) 18497 (Nigeria)	~400	4	2

Sequence features

The eight loci were sequenced from the individual sample in which they were identified, along with the human sample Hela and three other nonhuman primates, to test whether the recovered elements were specific to humans and, if so, to obtain the sequence of the ancestral preinsertion site of the *Alu* elements. In all cases, the *Alu* element was absent at the orthologous site in nonhuman primates, confirming the recent integration of the elements specifically in the human lineage. As shown in Fig. 1, all eight *Alu* elements displayed the hallmarks of recent integration, including conserved target site duplications ranging in size from 10 to 15 nucleotides and a long poly(A) tail at the 3′ end of the element ranging in size from 19 to 47 nucleotides [22,23].

Inspection of the nucleotide sequences of the eight *Alu* elements showed that two belong to the Ya8 subfamily (RC2 and RC3) while the remaining six elements belong to the Ya5 subfamily (Fig. 1). Although the protocol that we used is designed to preferentially recover Ya8 elements [22], the above results are not surprising for two reasons: (i) the Ya8 and Ya5 subfamilies are two closely related, human-specific *Alu* subfamilies [21] and (ii) the Ya8 subfamily comprises less than 50 copies in the human genome [11] while there are ~2000 copies of the Ya5 subfamily in the human genome [21]. Since all new loci that we identified are Ya8 or closely related Ya5 *Alu* elements, we believe that our approach is reasonably selective, especially when taking into consideration the fact that thousands of human-specific *Alu* insertions exist in the human genome [24]. This conclusion is further supported by the fact that the vast majority of the clones that we sequenced for each sample (e.g., ~78% of the 507 clones sequenced for sample L945) yielded either Ya8 or Ya5 *Alu* loci.

The eight *Alu* sequences that we identified in this study diverged from their respective subfamily consensus sequence by 0 to 2 nucleotide substitutions, further suggesting that they were integrated in the human genome very recently. Overall, all the sequence features associated with the eight *Alu* loci suggest that they may be recent enough to be highly polymorphic in human populations.

Population diversity

To test the degree of variation of the eight new *Alu* insertion loci in humans, we genotyped them in worldwide samples encompassing ~460 chromosomes (Table 2). All loci were found to be polymorphic, with the frequency of the allele with the *Alu* insert ranging from 1.4 to 45.0% and an average frequency across the eight loci of 23.0%. The frequencies recorded for the eight new loci were significantly lower (Mann–Whitney U test, $p=0.003$) than those recorded for the eight polymorphic *Alu*Ya8 loci (67.7% on average) previously

```
RC2       AGAACACTAC [AluYa8 / A19]     AGAACACTAC
RC3  AAAAAGTTAATTACA [AluYa8 / A5GA41] AAAAAGTTAATTACA
RC5   AAGATTTTTAAGTC [AluYa5 / A21]     AAGATTTTTAAGTC
RC6       AATACATGAGG [AluYa5 / A26]     AATACATGAGG
A1     AAAAAAATCACATC [AluYa5 / A39]     AAAAAAATCACATC
E1      AGAAAATAAGCTT [AluYa5 / A28GA3] AGAAAATAAGCTT
E2   AAAAACACAAAAAAG [AluYa5 / A28]     AAAAAMACAAAAAAG
E4      AATGTTACCCTGA [AluYa5 / A24]     AATTGTNCCCNTGN
```

Fig. 1. Sequence features of eight *Alu* insertion polymorphisms and their flanking sequence. *Alu* subfamily affiliation and length and content of the poly(A)-rich tail are indicated in brackets. Outside of brackets are nucleotide sequences immediately flanking the *Alu* insertions and corresponding to the target site duplications.

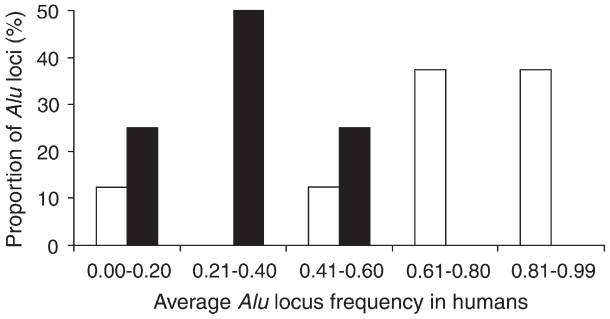

Fig. 2. Frequency distributions of polymorphic *Alu* elements. The frequency distribution of the eight *Alu* insertion polymorphisms identified in the study is shown in black, and that of the eight polymorphic *Alu* Ya8 elements identified in the human genome database [11] is shown in white. The *x* axis encompasses the different frequency classes of *Alu* insertion polymorphisms in the human population.

identified in the human genome reference sequence [11]. Furthermore, the frequency distribution of the eight new loci was clearly skewed toward low frequencies (Fig. 2). By contrast, the frequency distribution of the eight polymorphic *Alu*Ya8 loci previously identified in the human genome reference sequence [11] was skewed toward high insertion frequencies (Fig. 2). These results indicate that our approach preferentially detects low-frequency *Alu* elements in human populations, as compared to polymorphic *Alu* elements identified from genome database searches.

The average heterozygosity per locus was quite variable, with two extremely low values (~0.03 for RC5 and A1) and others that ranged from 0.30 to 0.46 (Table 2). While most loci were found at appreciable frequencies in all three major continental areas, two loci (RC5 and A1) displayed a remarkable distribution pattern since they were both found uniquely in populations with African ancestry. This pattern explains the extremely low heterozygosity values recorded for these loci on a worldwide scale, because all non-African individuals possessed the homozygous absent genotype at these two loci, thus leading to reduced genetic diversity on a global scale.

To further assess the restricted geographic distribution of the RC5 and A1 loci, we genotyped them in six additional populations from Africa, Asia, and Europe encompassing 280 different chromosomes (Table 3). The extended results were consistent with the results reported for the smaller dataset. In sum, the RC5 and A1 loci were genotyped in a worldwide panel of ~370 individuals (corresponding to ~740 different chromosomes) and they were both found exclusively in populations of African ancestry and completely lacking in non-African samples.

It is surprising that the two African-specific *Alu* loci are actually present in *every* African group (except RC5 in Egyptians, Table 2). This is because one would expect that *Alu* elements widespread in Africa would also be found outside Africa, as this would imply that the elements inserted in the human genome prior to the expansions of modern humans within and then out of Africa ~50,000 years ago [25,26]. By contrast, *Alu* loci restricted to Africa would be expected to have inserted in the human genome so recently that they would be found only in some but not all African populations. The fact that we found the two *Alu* loci RC5 and A1 in diverse African groups such as Pygmies, San, South African Bantus, African Americans (likely of West African ancestry), and Egyptians suggests that some migration that influenced all African groups may have

Table 2
Alu insertion frequencies for 12 human populations (for each locus, the frequency of the allele with the *Alu* insert is given) and average heterozygosity (Het) per locus

Population	Average *n*	RC2	RC3	RC5	RC6	A1	E1	E2	E4
African American	20	0.452	0.048	0.075	0.225	0.050	0.200	0.400	0.381
Pygmy	19	0.237	0.194	0.079	0.342	0.053	0.026	0.132	0.421
Egyptian	18	0.275	0.289	0	0.158	0.028	0.265	0.389	0.306
!Kung	19	0.211	0.368	0.029	0.175	0.026	0.100	0.368	0.658
AFRICA	76	0.297	0.221	0.048	0.224	0.039	0.145	0.322	0.442
East and Southeast Asia	20	0.550	0.250	0	0.175	0	0.025	0.725	0.175
South America	20	0.600	0.175	0	0.075	0	0.325	0.475	0.350
Indonesia	18	0.333	0.278	0	0.500	0	0	0.528	0.342
Pakistan	17	0.667	0.115	0	0.525	0	0.156	0.200	0.107
ASIA	75	0.529	0.211	0	0.319	0	0.128	0.481	0.253
Germany	20	0.571	0.211	0	0.119	0	0.250	0.425	0.150
Hungaria	18	0.556	0.167	0	0.026	0	0.333	0.500	0.167
Cyprus	20	0.425	0.225	0	0.250	0	0.425	0.400	0.300
France	19	0.575	0.175	0	0.200	0	0.375	0.316	0.250
EUROPE	77	0.532	0.195	0	0.150	0	0.346	0.409	0.216
WORLD	228	0.450	0.209	0.015	0.231	0.014	0.208	0.405	0.306
Average Het		0.464	0.325	0.029	0.319	0.026	0.297	0.451	0.389

Average *n*, average number of individuals genotyped for each locus.

Table 3
Alu insertion frequencies for six human populations (for each locus, the frequency of the allele with the *Alu* insert is given)

Continental region	Population	Average *n*	RC5	A1
Africa	Sotho-Tswana	48	0.010	0.106
	Nguni	45	0.011	0.211
Asia	Bangladeshi	16	0	0
	Indonesia	12	0	0
Europe	Swiss	7	0	0
	Bretons	12	0	0

Average *n*, average number of individuals genotyped for each locus.

taken place after modern humans left Africa ~50,000 years ago. However, further studies involving more loci are needed to test whether such a migration indeed occurred or whether the observed pattern based on two *Alu* loci is the result of chance alone.

Concluding remarks

In sum, our approach was able to recover two new *Alu* elements that exhibit restricted geographic distributions. The two African-specific *Alu* loci that we identified may prove useful for human evolution studies or forensics applications. It is noteworthy that these two loci were found within African populations at low frequencies (<25%). Therefore, the absence of the element in an individual would not be informative with respect to geographic affinities. However, the presence of the element in an individual would suggest an African ancestry with a high probability. Ideally, several *Alu* markers with relevant geographic distributions could be used in conjunction to increase the resolution and confidence in the results. Using the available data, we can estimate how many *Alu* loci would be needed to infer geographic affiliation with a high degree of confidence. Assuming that the loci are in Hardy–Weinberg equilibrium and in linkage equilibrium and assuming that they are truly African-specific, then the presence of at least one *Alu* insert at any locus indicates that the individual is of African ancestry. So, individuals from Africa are incorrectly classified only if they are homozygous for the absence of the insert at both loci. The two African-specific loci RC5 and A1 have an average insert frequency of 0.066 in African populations, so the frequency of homozygotes for the absence of the insert at both loci is $(0.934)^2 \times (0.934)^2 = 0.761$, and the probability of correctly classifying an African individual as African is 1-0.761 = 0.239 based on two loci. In general, for *n* loci with an average insert frequency of *p*, the probability of correctly classifying an individual is $1-((1-p)^2)^n$). If the average insert frequency is 0.066, then 22 loci would be needed to have a 95% chance of correctly classifying an individual as African. In other words, with 22 African-specific loci, each with average insert frequency of 0.066, there is less than a 5% chance that an African individual would be homozygous for the absence of the insert at all 22 loci. Thus, it may ultimately become possible to correctly infer the geographic affiliation of unknown samples with high levels of confidence without having to genotype as many as 100 *Alu* loci [17]. To this end, the identification of additional informative *Alu* loci is desirable, and we demonstrated here that our approach is capable of recovering new *Alu* loci with restricted geographic distribution.

Materials and methods

DNA samples

A total of 17 human DNA samples were used to ascertain new *Alu* insertions, either separately or as pools of three individuals (Table 1). DNA samples were obtained from the Coriell Institute for Medical Research, except L945, 10237, 10408, and 1052, which were available from previous studies in our laboratory. The human-specific nature of the candidate *Alu* loci was evaluated by their presence or absence in four primate species, including human HeLa (cell line ATCC-CCL2), common chimpanzee [Clint] (NS06006B), gorilla (AG05251), and orangutan (ATCC-CR6301). Each locus was also genotyped in a panel of 228 individuals (456 chromosomes) from 12 human populations originating from the three major continental groups: Africa, Asia, and Europe [12,13]. Details on the populations and sample sizes are shown in Table 2. We assigned our South American samples to the Asian continental region because of the genetic roots of Amerindians in Asia [27]. *Alu* loci RC5 and A1 were further genotyped in 140 additional individuals (280 chromosomes) from six diverse human populations (Table 3) [12,13].

Identification of candidate Alu insertion loci

We used a modification of the ASAP PCR previously described [22]. Genomic DNA was digested with *Nde*I. Double-stranded linkers MSET and MSEB [22] were subsequently ligated to the digested DNA. Three successive PCRs were then performed using the linker primer LNP and nested *Alu* primers ASII, HS18R, and HS16R [22] to obtain collections of PCR products enriched in *Alu* elements belonging to the Ya8 subfamily. The *Alu*Ya8 subfamily is one of the youngest and most polymorphic *Alu* subfamilies currently known in humans [11,22]. Third-round PCR products were cloned into vectors using the TOPO-

Table 4
Candidate *Alu* loci amenable to PCR and PCR amplification conditions

Locus name	Genomic location	Forward primer (5′>3′)	Reverse primer (5′>3′)	Annealing temperature	PCR product size (filled/empty site)
RC2	8p11.21	TCATCTTGACCTTGCAGCAC	CCGTAACAGGACAGCTCACA	60	420/120
RC3	3p24.1	TGCCCAAATGTAGACCTTGT	TTGTTGGAGTTGAGGCATCTT	59	460/160
RC5	7q31.31	CCCCTTCCAGAGAAGCATTT	GCTCTTCTTTTTCAGTGAGTTTCC	60	460/160
RC6	12q23.3	CATATGCACCGCGCTAAGTA	TCCTAATGCCTTTTTCCATAACA	60	470/170
A1	14q31.3	TTGAAAAGGGGGTAGTTTATGA	AACCCTTGAGAGGGGATGAC	55	710/384
E1	15q25.2	GCCCTCTAGGAAAGTGAAAGAA	GCCATAGGTTTCCTCTGGTTG	55	570/242
E2	13q12.3	CCAAGACCCAGGCATTAAAA	GAAGGATGGTGATTGCAGGT	55	523/202
E4	5q14.3	TATGCATAGCCAAAAGAGAGCA	ACATTTGGGACATCAGGGTAAC	60	497/197

TA cloning kit (Invitrogen), according to the manufacturer's instructions. Colonies were randomly picked and DNA sequencing was performed using chain termination sequencing on an Applied Biosystems 3100 automated DNA sequencer.

The resulting sequences contained the 5′ region of an *Alu* element along with the *Alu* element 5′ flanking sequence. The flanking sequences were used as queries in BLAT searches against the May 2004 freeze of the human genome reference sequence, as implemented in the University of California, Santa Cruz, genome browser (http://genome.cse.ucsc.edu) to determine whether an *Alu* insertion was already known to be located at each locus.

Genotyping of Alu loci

When the BLAT search predicted the absence of the *Alu* element in the human genome reference sequence, 1000 bp of flanking sequence from each side of the predicted *Alu* insertion site were extracted and oligonucleotide primers were designed as previously described [11]. PCRs were conducted to amplify the candidate loci first in four nonhuman primate species and then in various human samples, as previously described [11]. Specific information on each locus including chromosomal location, primer sequences, annealing temperature, and PCR product sizes is shown in Table 4. Resulting PCR products were separated on 2% agarose gels, stained with ethidium bromide, and visualized using UV fluorescence. PCR products from the four nonhuman primate samples were sequenced as described above. Sequences generated in this study have been deposited in GenBank under Accession Nos. EF372292–EF372328.

Acknowledgments

We thank K. Han for technical assistance and A.-H. Salem, S. Milligan, J.-P. Moisan, M. Hochmeister, L. Henke, J. Henke, M. Tahir, and P. Ioannou for providing samples. This research was supported by National Science Foundation BCS-0218338 and EPS-0346411 (MAB), National Institutes of Health RO1 GM59290 (MAB), and the State of Louisiana Board of Regents Support Fund (MAB).

References

[1] M.A. Batzer, P.L. Deininger, Alu repeats and human genomic diversity, Nat. Rev. Genet. 3 (2002) 370–379.
[2] P.L. Deininger, M.A. Batzer, C.A. Hutchison III, M.H. Edgell, Master genes in mammalian repetitive DNA amplification, Trends Genet. 8 (1992) 307–311.
[3] R. Cordaux, D.J. Hedges, M.A. Batzer, Retrotransposition of Alu elements: how many sources? Trends Genet. 20 (2004) 464–467.
[4] R. Cordaux, D.J. Hedges, S.W. Herke, M.A. Batzer, Estimating the retrotransposition rate of human Alu elements, Gene 373 (2006) 134–137.
[5] E.S. Lander, et al., Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921.
[6] J.M. Chen, P.D. Stenson, D.N. Cooper, C. Ferec, A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease, Hum. Genet. 117 (2005) 411–427.
[7] S.K. Sen, et al., Human genomic deletions mediated by recombination between Alu elements, Am. J. Hum. Genet. 79 (2006) 41–53.
[8] P.A. Callinan, et al., Alu Retrotransposition-mediated Deletion, J. Mol. Biol. 348 (2005) 791–800.
[9] J.A. Bailey, G. Liu, E.E. Eichler, An Alu transposition model for the origin and expansion of human segmental duplications, Am. J. Hum. Genet. 73 (2003) 823–834.
[10] J. Wang, et al., dbRIP: a highly integrated database of retrotransposon insertion polymorphisms in humans, Hum. Mutat. 27 (2006) 323–329.
[11] R. Cordaux, J. Lee, L. Dinoso, M.A. Batzer, Recently integrated Alu retrotransposons are essentially neutral residents of the human genome, Gene 373 (2006) 138–144.
[12] M.A. Batzer, et al., African origin of human-specific polymorphic Alu insertions, Proc. Natl. Acad. Sci. USA 91 (1994) 12288–12292.
[13] M. Stoneking, et al., Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa, Genome Res. 7 (1997) 1061–1071.
[14] M.A. Batzer, et al., Genetic variation of recent Alu insertions in human populations, J. Mol. Evol. 42 (1996) 22–29.
[15] W.S. Watkins, et al., Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms, Genome Res. 13 (2003) 1607–1618.
[16] M.J. Bamshad, et al., Human population genetic structure and inference of group membership, Am. J. Hum. Genet. 72 (2003) 578–789.
[17] D.A. Ray, et al., Inference of human geographic origins using Alu insertion polymorphisms, Forensic Sci. Int. 153 (2005) 117–124.
[18] A.B. Carter, et al., Genome-wide analysis of the human Alu Yb-lineage, Hum. Genomics 1 (2004) 167–178.
[19] J. Wang, et al., Whole genome computational comparative genomics: A fruitful approach for ascertaining Alu insertion polymorphisms, Gene 365 (2006) 11–20.
[20] E.A. Bennett, L.E. Coleman, C. Tsui, W.S. Pittard, S.E. Devine, Natural genetic variation caused by transposable elements in humans, Genetics 168 (2004) 933–951.
[21] A.C. Otieno, et al., Analysis of the Human Alu Ya-lineage, J. Mol. Biol. 342 (2004) 109–118.
[22] A.M. Roy, et al., Recently integrated human Alu repeats: finding needles in the haystack, Genetica 107 (1999) 149–161.
[23] A.M. Roy-Engel, et al., Active Alu element "A-tails": size does matter, Genome Res. 12 (2002) 1333–1344.
[24] Chimpanzee Sequencing and Analysis Consortium, Initial sequence of the chimpanzee genome and comparison with the human genome, Nature 437 (2005) 69–87.
[25] R.L. Cann, M. Stoneking, A.C. Wilson, Mitochondrial DNA and human evolution, Nature 325 (1987) 31–36.
[26] P.A. Underhill, et al., The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations, Ann. Hum. Genet. 65 (2001) 43–62.
[27] C.J. Mulligan, K. Hunley, S. Cole, J.C. Long, Population genetics, history, and health patterns in native americans, Annu. Rev. Genomics Hum. Genet. 5 (2004) 295–315.