 Int J Hum Genet, 8(1-2): 61-74 (2008)

Role of Alu Element in Detecting Population Diversity

Manorama Tripathi[1,2], U.K. Chauhan[2], Piyush Tripathi[1] and Suraksha Agrawal[1]

1. Department of Medical Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Raebareli Road, Lucknow 226 014, Uttar Pradesh, India
2. Department of Bitotechnology, APS Univeristy, Rewa, Madhya Pradesh, India

KEYWORDS Alu insertion; genetic diversity; North Indian populations; Caste system; human evolution

ABSTRACT Indel polymorphisms are becoming increasingly useful markers for population genetic studies. These are stable mutational events that are unlikely to undergo reverse or convergent mutations, making them useful as markers for distinguishing chromosomal lineages identical by descent and for using in population diversity studies. There has been a great deal of interest to use retrotransposones like Alu element, which displays indel polymorphism, in the population diversity studies. The polymorphic Alu insertions are particularly useful for such studies because the probability of independent retroposition at the same exact chromosomal site is virtually zero. All loci carrying a particular Alu insertion are derived from a unique event and hence are identical by descent. Further the ancestral state for polymorphic Alu insertions is the absence of the insertion hence the direction of mutational change is the gain of the Alu element at a particular locus. Knowing the ancestral state and the direction of mutational change greatly facilitates the analysis of population relationship but is generally not possible for other types of loci. These features make Alu insertion polymorphisms more attractive than other autosomal classical markers for analyses of population history and structure. This review is focused on Alu element and its role in detecting population diversity with few empirical examples from around the world. The phylogenetic studies based on Alu polymorphisms on Indian populations has also been presented.

INTRODUCTION

When and where our species originated remains the curious question, but archeological and genetic evidences has thrown light upon it. History of human species also has left genetic signals in regional populations. For example, populations in Africa have higher levels of genetic diversity and tend to have lower amount of linkage disequilibrium than do populations outside Africa, partly because of the larger size of human populations in Africa over the course of human history and partly because the number of modern humans who left Africa to colonize the rest of the world appears to have been relatively low (Gabriel et al. 2002). The studies of mitochondrial DNA, Alu polymorphisms, and other genetic markers have significant implications in deciphering human ancestry. Genetic data can be used to infer population structure and assign individuals to groups that often correspond with their self-identified geographical ancestry. However, in the analysis that assigns individuals to groups it becomes less apparent that self-described racial groups are reliable indicators of ancestry. One cause of the reduced power of the assignment of individuals to groups is admixture. Some racial or ethnic groups do not have homogenous ancestry. African Americans have a mixture of West African and European ancestry. It was demonstrated that on an average African Americans have ~80% African ancestry (Shriver et al. 2004). Similarly, many white Americans have mixed European and African ancestry, where ~30% of whites have less than 90% European ancestry.

Corresponding Author: Prof. Suraksha Agrawal
Department of Medical Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Raebareli Road, Lucknow 226014, Uttar Pradesh, India
Telephone: 091-522 -2668004-8 Ext 2338, 2346, 2347
Fax: 091-522 -2668973/2668017
E-mail: suraksha@sgpgi.ac.in

INDIAN PERSPECTIVE

Indian sub-continent consists of more than 1 billion humans constituting ~1/5th of the total world population. Four different morphological groups, namely Negritos (Andaman Island, Nilgiri hills of Tamil Nadu), Australoids (Central and southern region), Mongoloids (sub-Himalayan and north eastern region) and Caucasoids (spreaded all over India, especially north and north-western regions) are found in India. Linguistically, there are two most frequent language families, Elamo-Dravidian, spoken mostly in southern India and Indo-European,

spoken almost all over India. The two families are structured into 18 official languages and ~750 dialects. Apart from these, there are Austro-Asiatic speakers, which are mainly tribes and Sino-Tibetan speakers found exclusively in northeastern part of the country.

Various unique features of population structure draw the attention of investigators to explore the gene pool of Indian populations. First is its crucial geographical location that lies on the postulated southern coastal route followed by the anatomically modern *H. sapiens* while coming out of Africa (Cann 2001; Lahr and Foley 1998). Second is the extensive gene flow through a series of migrations and invasions that have contributed the contemporary genetic variation across different geographical locations of India and have created enormous genetic diversity. The 'Arab-Indian' haplotype at the α-globin gene cluster is a strong testimony to the movements of genes with people. First gene flow to India was witnessed about 10-15 KYA, with the agricultural development in the Fertile Crescent region, as a part of an eastward wave of human migration (Cavalli-Sforza et al. 1994; Renfrew 1989). This wave brought Dravidian languages, mainly Elamo-Dravidian languages (Ruhlen 1991), into India (Renfrew 1989), which may have originated in the Elam province (Zagros Mountains, southwestern Iran) and are now confined to southern India and to some isolated groups in Pakistan and northern India. Further migration event was the arrival of pastoral nomads from the central Asian steppes to the Iranian plateau, ~4,000 YBP that carried Indo-European language family, which eventually replaced Dravidian languages from most of Pakistan and northern India, perhaps by an elite dominance process (Renfrew 1989; Quintana-Murci et al. 2001). These nomadic migrants became strengthened by admixing with native Dravidic-speaking proto-Asian populations and subsequently established the Hindu caste hierarchy to legitimize and maintain their power (Cavalli-Sforza et al. 1994).

The people of India are culturally stratified as tribals and nontribals. The most unique feature of the population structure of India is the caste system that has sub structured Indian population in various endogamous groups. Ethnographic and genetic evidence both support that Hindu castes have been highly endogamous for a long time (Bamshad et al. 2001; Misra 2001). Although the level of genetic differentiation between castes is relatively small, genetic distances observed in several studies suggest that gene flow is limited (Papiha et al. 1996; Bamshad et al. 2001; Dutta et al. 2002; Lakshmi et al. 2002).

MARKERS CHOSEN FOR GENETIC ANALYSIS

The distribution of genetic variation within and among human populations has long been an important tool for inferring the evolutionary history of modern humans. Numerous early investigators noted that human groups differ in skin color, hair form, body size, and other physical characteristics. Subsequently, blood and protein markers and other traditional genetic polymorphisms have been used to study the extent of variation and relationships among the human populations, both at the local and global scales (Cavalli-Sforza et al. 1994). Dramatic improvements in genotyping technologies over the past 3 decades and the emergence of PCR have facilitated the development of many types of DNA markers. Depending upon the type these could be STRs, SNPs and indels, where as based on their chromosomal origin these could be autosomal, mitochondrial and Y-Chromosome .

In recent years, considerable attention has been devoted to both uniparental and autosomal genetic markers. Because of their lack of recombination, uniparental markers, viz., mtDNA and the nonrecombining region of the Y chromosome facilitate easier construction of haplotypes that help in tracing maternal and paternal lineages. This is not the case with the recombining markers (Ingman et al. 2000; Underhill et al. 2000). Nevertheless, though recombination in case of autosomal loci may mislead about their history, simultaneous use of an array of autosomal polymorphic loci spread across the genome provides more general inference about demographic history and population relationships compared to gender specific markers of Y chromosome and mtDNA.

Studies of autosomal variation that are based on protein polymorphisms, blood groups, restriction-site polymorphisms, and Alu insertions have revealed much about within- and among-population genetic diversity of humans (Cavalli-Sforza et al. 1994; Relethford 2001). Among autosomal markers, special attention has been paid to polymorphisms of short tandem repeats (STRs). These loci are numerous, highly

polymorphic, and densely distributed across the genome, and they mutate at a high rate approaching 10^{-3} per generation (Weber and Wong, 1993), facilitating inferences about short-term evolution. Microsatellite statistics have been exploited to study population antiquity, expansion (Jin et al. 2000; Zhivotovsky et al. 2001) and migration (Rousset, 1996). However, a large number of such markers are required in order to get reliable conclusions (Goldstein et al. 1996; Jorde et al. 1997; Zhivotovsky et al. 2001). SNPs are biallelic markers. These polymorphisms, which are expected to occur in the human genome every few kilobases, are abundant and are thought to have low mutation rates. Nucleotide variability can be detected using several methods including SSCP, dHPLC and by direct sequencing. Although emerging SNP-based technologies have recently proven quite useful, SNPs have some limitations due to the fact that they represent single base pair differences. Loci that are highly informative and heterozygous are accompanied by a high mutation rate where as loci that have lower heterozygosities are accompanied by a lower mutation rate. The specific choice will depend on the nature of a given study. Like most other genetic polymorphisms, SNPs can be merely identical-by-state; that is, they may have arisen as a result of an independent parallel forward or backward mutation resulting in genotype misclassification (homoplasy). Recent improvements in SNP-based approaches suggest that most of these problems can be overcome by carefully selecting the correct loci and the correct number of SNPs to be used.

Indel polymorphisms are becoming increasingly useful markers for population genetic studies. It is estimated that in/dels comprise approximately 8% of all human polymorphisms. Insertion or deletions can range in size from 1 nucleotide to several million nucleotides. These are stable mutational events that are unlikely to undergo reverse or convergent mutations, making them useful as markers for distinguishing chromosomal lineages identical by descent and for using gene mapping studies.

There are two reasons why these polymorphic Alu insertions should be particularly useful for population genetic studies: 1. the probability of independent retroposition at the same exact chromosomal site is virtually zero; 2. all loci carrying a particular Alu insertion are derived from a unique event and hence are identical by descent (Stoneking et al. 1997). Polymorphic Alu insertions should thus more accurately reflect population relationship than markers such as RFLPs and VNTRs in which the sharing of the same allele by two individual may reflect chance identity by state (independent mutations). The ancestral state for polymorphic Alu insertions can be reasonably inferred to be the absence of the insertion, and the direction of mutational change is therefore the gain of the Alu element at a particular locus (Batzer et al. 1996; Stoneking et al. 1997). Knowing the ancestral state and the direction of mutational change greatly facilitates the analysis of population relationship which is generally not possible for other types of loci. These features make Alu insertion polymorphisms more useful than other autosomal markers (such as classical markers) for analyses of population history and structure.

Alu Elements

Recent genome sequencing projects have concluded that the non-coding DNA occupies most of the genome compared to actual coding DNA in eukaryotic genomes (Lander et al. 2001; Waterston et al. 2002). The non coding region is also called as junk DNA. Majority of the 'junk' DNA consists of transposable elements. These elements contribute ~half of the total length of the human genome (Lander et al. 2001). Recent studies of active transposable elements suggest that they can alter gene expression with either harmful or beneficial effects, especially when an active gene is targeted (Deininger and Batzer, 1999; Mills et al. 2007). These elements are associated with genomic fluidity via *de novo* insertions, insertion-mediated deletions, and recombination events (Deininger and Batzer 1999; Gilbert et al. 2002; Symer et al. 2002; Callinan et al. 2005; Han et al. 2005; Sen et al. 2006). The picture now emerging from the literature demonstrates that these transposable elements have various functions and play a significant and dynamic role in the process of the evolution of genomes.

These transposable elements were first discovered by Barbara McClintock in the variegated color pattern in maize (McClintock 1956). In mammals, transposable elements are broadly divided into two categories, transposons and retrotransposons, based on their nature of mobilization. DNA transposons move in a "cut and paste" mechanism and are currently active in various genomes like bacteria, plants, and

insects (Mizuuchi 1992; Lander et al. 2001). However, retrotransposons mobilize to a new location in the genome via an RNA intermediate, thereby duplicating the element (Luan et al. 1993; Feng et al. 1996; Moran et al. 1996). As a result of this "copy and paste" mechanism, retro-transposons accumulate much faster in the genome and have a major impact on genomic architecture (Deininger and Batzer 2002).

Discovery and Structure of Alu Elements

Alu elements are repetitive elements. These were identified originally almost 35 years ago as a component in human DNA renaturation curve (Houck et al. 1979). As these sequences included Alu I restriction sites they were named as Alu elements (Houck et al. 1979). Detailed analysis of this portion of the renaturation curves enabled the sequence analysis of individual Alu elements.

Full-length Alu elements are ~300 bp long followed by a tail of 20-30 adenosine bases. It has dimeric structure composed of two similar but distinct monomers, linked with an oligo (dA) tract (Fig. 1). The current organization is suggestive of their formation due to ancient duplication, as two poly 'A' stretches are present in the elements. The right Alu monomer (3' half) contains a 31 bp insert, which is absent from the left monomer (5' half). The left monomer (5' half) of each sequence contains an RNA polymerase III promoter having two functional boxes (A and B box) but is absent from the right monomer (Fuhrman et al. 1981).

Alu sequences do not encode d (T) 4 RNA polymerase III transcription terminator signal, one often present in the flanking down stream genomic sequence. At the 3' end of Alu element contains a poly (A) tail of variable length. Newly retro transposed Alu elements are flanked by short intact direct repeats that are derived from the site of insertion, which is consistent with the insertion into staggered nicks in the DNA (Hutchinson et al. 1993).

Occurrence and Distribution of Alu Sequences

Alu elements are commonly found in introns, 3' untranslated region of genes and intergenic genomic regions. Alu sequences accumulate preferentially in gene rich regions. The long-lasting presence of Alu elements along with their high copy number in the genome raises questions regarding their role during evolution.

Several potential functions have been proposed for Alu elements, such as stimulation of protein synthesis under stress conditions and regulation of gene expression, but later on these could not be confirmed (Deininger and Batzer 1999). Recent evidences suggested their non uniform distribution across the human genome, as older Alu elements are preferentially found in GC-rich regions while younger Alu elements are slightly more abundant in AT-rich regions (Lander et al. 2001). This was suggestive of positive selection to maintain them in GC-rich, gene-rich regions (Lander et al. 2001). However, this assumption also could not hold true regarding the time scale on which the selective process might have acted (Brookfield 2001; Batzer and Deininger 2002).

Recently the genomic distribution of the three youngest human Alu subfamilies (Ya5a2, Ya8 and Yb9) in conjunction with their insertion polymorphism status in the human population was analyzed, since selection can only act on polymorphic elements (Cordaux et al. 2006). Results indicated that: (i) polymorphic and fixed recently integrated Alu elements are found in

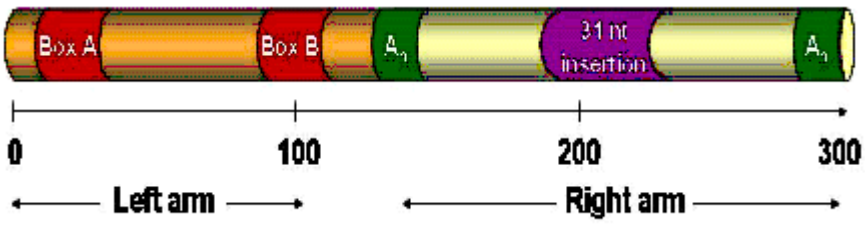

Fig. 1. Architecture of Alu elements.
The dimeric structure of Alu elements composed of two monomers, each followed by a short poly(A) tail.

genomic regions whose GC contents are statistically indistinguishable, and (ii) recently integrated Alu elements are inserted randomly, regardless of the GC content of the surrounding genomic DNA. These evidences show that recently integrated "young" Alu elements are not subject to positive or negative selection at larger scale. Therefore, young Alu elements can be regarded as essentially neutral residents of the human genome (Cordaux et al. 2006). These results also imply that selective processes specifically targeting Alu elements can be ruled out as explanations for the accumulation of Alu elements in GC-rich regions of the human genome.

Origin and Amplification of Alu Elements

The origin and amplification of Alu elements are evolutionarily recent which occurred with the radiation of primates in the past 65 million years (Deininger and Daniels, 1986). Sequence analysis of Alu RNAs has indicated that these were ancestrally derived from 7SL RNA gene, which forms a part of the ribosome complex (Ullu and Tschudi, 1984) and also an abundant cytoplasmic component of the signal recognition particle that mediates the translocation of secreted proteins across the endoplasmic reticulum (Okada 1991). Initially it was estimated that these mobile elements are formed in high copy number of ~ 500,000 copies comprising approx 5% of the genome (Batzer et al. 1996) with an average spacing of 4 kb. Detailed analysis of data from HGP has confirmed that they have more than one million copies, constituting the most abundant mobile element in the human genome comprising ~ 10% of the human genome (Batzer and Deininger 2002). Therefore, the origins of present day Alu elements dispersed throughout the human genome can be traced to initial gene duplication during primate evolution and their subsequent amplification. This type of duplication, followed by the expansion of a SINE family, has occurred sporadically throughout evolutionary history in mammalian and non-mammalian genomes (Deininger and Batzer 2002).

Alu elements are retroposons i.e. they transpose via an RNA intermediate, which is then reverse-transcribed and inserted into a new location in the genome. This RNA-mediated transposition is termed retroposition (Rogers

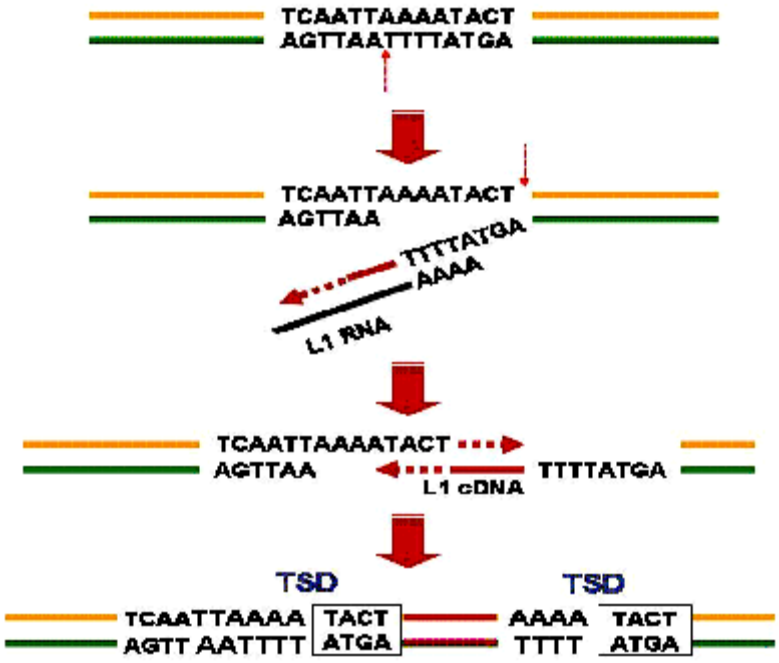

Fig. 2. Target-primed reverse transcription (TPRT) mechanism.

1998). The most accepted mechanism for Alu mobilization is shown in Figure 2. The Alu-derived transcript uses a nick at its genomic integration site for target- primed reverse transcription (TPRT) to occur (Feng et al. 1996; Moran et al. 1996; Luan et al. 1993). As the Alu element does not code for an RNA-polymerase-III termination signal, its transcript will therefore extend into the flanking unique sequence. The typical RNA-polymerase-III terminator signal is a run of four or more Ts on the sense strand, which results in three Us at the 3' terminus of most transcripts. It has been proposed that the run of A at the 3' end of the Alu might anneal directly at the site of integration in the genome for target-primed reverse transcription. It seems likely that the first nick at the site of insertion is often made by the L1 endonuclease at the TTAAAA consensus site. The mechanism for making the second-site nick on the other strand and integrating the other end of the Alu element remains unclear. A new set of direct repeats is created during the insertion of new Alu element.

The L1 endonuclease cleaves between an A and a T on the bottom strand of DNA, exposing a 3' hydroxy group. The target DNA functions as a primer from which complementary DNA is made (using the polyadenylated L1 RNA as a template). The second cleavage (on the sense strand of the chromosomal DNA) is staggered relative to the first, and repair or filling in of the gaps between the L1 cDNA and the chromosomal DNA results in a short target-site duplication (TACT, boxed).

Alu Polymorphism and Human Genetic Diversity

The evolution of modern human populations continues to be a topic of controversy. Evidence from mtDNA, Ychromosome polymorphisms, autosomal markers, and fossil material supports both the expansion of early modern human populations in Africa and the partial or complete replacement of other hominid groups (Ovchinnikov et al. 2000). Interpretation of other, primarily fossil data suggests both an early expansion of hominid lines and multiregional development of modern humans (Wolpoff et al. 2001). The increasing number and variety of genetic markers offer additional opportunities for more-detailed analysis of human evolution and of genetic diversity within and between human populations.

Numerous studies utilizing a variety of polymorphic loci suggest an overall pattern of higher gene diversity in African populations compared to that in the non-African populations (Jorde et al. 2000). These studies have focused primarily on neutral restriction-site polymorphisms (RSPs), short tandem-repeat polymorphisms (STRPs), noncoding autosomal sequences, Y chromosomes, and mtDNA. Analyses of protein coding regions, including the neurofibromatosis type 1 (NF1), angiotensin-converting enzyme (ACE), myotonic dystrophy (DM), dopamine D2 receptor (DRD2) and fragile X (FMR1) loci, have also shown higher levels of diversity in African populations compared with levels found in non-African populations (Rieder et al. 1999; Crawford et al. 2000). Some loci, including the melanocortin 1 receptor (MCR1) and phenylalanine hydroxylase (PAH) loci, do not consistently show patterns of higher diversity in African populations, revealing the potential influence of natural selection on patterns of genetic diversity. Noncoding DNA sequences on chromosomes 22, 15, and 1 show higher nucleotide-diversity estimates for African populations than for non-African populations, consistent with a recent human population expansion (Yu et al. 2001).

Alu insertion polymorphisms and other SINE elements are robust markers for evolutionary and phylogenetic studies because they have a unique mutational mechanism, an absence of back mutation, and a lack of recurrent forward mutation (Hamdi et al. 1999; Roy-Engel et al. 2001). A specific Alu insertion and nearby flanking sequence will be identical by descent in all individuals in whom they occur (Batzer et al. 1996). Thus, sets of related chromosome regions marked by an Alu insertion event can be distinguished from a pool of ancestral chromosomes that lack the element. These features give each locus genetic polarity that allows the independent assignment of an ancestral state and a root for phylogenetic analyses.

Previous studies of human genetic variation have utilized polymorphic Alu insertions to gain insight into population history. Studies using multiple Alu loci or a single Alu locus with flanking markers show high African diversity and a greater effective population size for Africans (Stoneking et al. 1997; Watkins et al. 2002). When a large number of Alu elements are analyzed, individuals can usually be classified according to their continent of origin (Bamshad et al. 2003). Alu

insertions are also useful for resolving genetic relationships in more limited locales such as NW Africa and the Caucasus region (Comas et al. 1998).

The use of these polymorphisms in a world-wide survey of human populations has confirmed the African origin of modern humans (Batzer et al. 1996; Stoneking et al. 1997). However, the use of Alu insertion polymorphisms in human evolution has been focused world-wide, and except for some population studies, relatively little research has been devoted to specific population questions. Comas et al. 2000 analyzed 11 Alu insertion polymorphisms (ACE, TPA25, PV92, APO, FXIIIB, D1, A25, B65, HS2.43, HS3.23, and HS4.65) in several NW African and Iberian populations.

Watkins et al. in 2002 have examined 35 Alu loci and 30 gene-related RSPs in 31 world populations, to characterize diversity and genetic structure in modern human populations. The genetic-distance estimates based on Alu-insertion polymorphisms, RSPs, STRPs, and mtDNA, using the same individuals in the same populations, are highly correlated. They have demonstrated that the African populations have higher overall Alu gene diversity (heterozygosity) than do other populations. This finding is concordant with results from other marker systems, including the mitochondrial genome, autosomal STRPs, and the Y chromosome (Jorde et al. 2000). The observed homozygosity for the Alu insert (the derived state) is significantly higher ($P<0.05$) in all non-African groups than it is in sub-Saharan Africans. All subpopulations outside Africa have higher observed levels of Alu homozygosity, as averaged over the 35 Alu loci, compared to populations within Africa. Comparisons of the major population groups and their subpopulations, through the use of rooted RSP loci, show a similar trend, outside Africa, toward increasing frequency and homozygosity of derived alleles.

This trend was similar to Stoneking et al. 1997, where they have investigated 34 independent world populations using only eight Alu loci. They have also confirmed that the average Alu-insertion frequencies over all loci show a similar trend toward an increasing frequency of Alu insertions in populations outside Africa.

Nasidze et al. (2001) analysis of 8 Alu insertion loci (ACE, TPA25, PV92, APO, FXIIIB, D1, A25, B65) in six populations from the Caucasus exhibits low levels of within-population variation and high levels of between-population differentiation, with the average F_{ST} value for the Caucasus of 0.113, which is almost as large as the F_{ST} value of 0.157 for worldwide populations (Table 1).

Table 1: Average heterozygosity and F_{ST} values for various populations.

Region	*No. of populations*	*No. of individuals*	*Heterozygosity*	F_{ST}
Africa	6	176	0.402	0.086
Americas	4	184	0.381	0.038
Europe	7	334	0.396	0.017
Sahul	3	185	0.308	0.105
Southeast Asia	7	359	0.377	0.067
Western Asia	7	262	0.414	0.053
Caucasus	6	221	0.311	0.113
World	40	1721	0.426	0.157

(Nasidze et al., *European Journal of Human Genetics* 2001, 9, 267 - 272)

The average heterozygosity for each locus ranged from 0.09 to 0.47, which is close to the maximum possible heterozygosity of 0.5 for a bi-allelic locus. The average heterozygosity for each population was lower in Ingushians (0.209) than in the other populations (range 0.318 ± 0.394). The Fst values are also substantial (Tables), ranging from 0.018 (APO and A25) to 0.222 (FXIIIB). The average Fst value for the Caucasus is almost seven times higher than in Europe and twice to that in west Asia (Table 1). This high Fst value indicates large differences in the Alu insertion frequencies among Caucasus populations. The Caucasus groups were compared with a worldwide data set typed for the same eight Alu insertion loci (Stoneking et al. 1997) to examine population relationships. Neither geographic nor linguistic relationships appear to explain the genetic relationships of Caucasus populations. Instead, it appears as if they have been small and relatively isolated, and hence genetic drift has been the dominant influence on the genetic structure of Caucasus populations.

Bamshad et al. (2003) demonstrated that, for a collection of heterogeneous samples from sub-Saharan Africa, East Asia, and Europe, the genetic data accurately predicted assignment to clusters that corresponded to major continents. However, correct assignment to the continent of origin with a mean accuracy of at least 90% required a minimum of ~60 Alu markers or microsatellites. This is a modest number of markers, but it supports the contention that most studies performed to date have lacked the power to make

strong inferences about population structure and sample assignment, even among highly differentiated samples (Wilson et al. 2001; Romualdi et al. 2002). One criterion that has been used to rank the power of loci for detecting population structure is F_{ST} (Bowcock et al. 1991; Rosenberg et al. 2001). For some Alu markers, the insertion frequency varied little among continental populations, whereas others were nearly monomorphic in one continental population or another. Accordingly, the F_{ST} estimates of individual Alu loci ranged from 0 to 0.72. Thus, although a minimum of 60 Alu markers or microsatellites were necessary to assign the predicted continent of origin for at least 90% of all the samples, individual markers were not equally informative. The F_{ST} value among the continental populations was markedly lower for microsatellites (0.042) than for Alu markers (0.13). The Alu F_{ST} value is consistent with F_{ST} estimates obtained in previous studies of biallelic markers.

One of the biggest data set is contributed by Watkins et al. 2002, where they have used 100 Alu polymorphic loci scattered over all 22 autosomes to investigate 710 individuals representing 31 populations from Africa, East Asia, Europe, and India (Table 2). They have demonstrated that the Alu diversity is highest in Africans (0.349) and lowest in Europeans (0.297). The diversity trends using 100 Alu insertion polymorphisms are, in general, consistent with studies using smaller numbers of Alu loci (Stoneking et al. 1997; Watkins et al. 2002). Alu insertion frequency is lowest in Africans (0.463) and higher in Indians (0.544), E. Asians (0.557), and Europeans (0.559).

In this study F_{ST} results suggested that Alu diversity between populations is highest for continental groups that are separated by large geographic distances, such as Africa, Asia, and Europe. The reduction in F_{ST} observed when Indians are included is consistent with previous work showing both Asian and European affinities in South Indian populations (Bamshad et al. 2001). Large genetic distances are observed among African populations and between African and non-African populations. The root of a neighbor-joining network is located closest to the African populations. These findings are consistent with an African origin of modern humans and with a bottleneck effect in the human populations that left Africa to colonize the rest of the world.

Cotrim et al. in 2004 studied four Alu polymorphic loci (APO, ACE, TPA25, and FXIIIB) among individuals from six Brazilian African-Derived Populations. The allelic frequencies in these populations were similar to those previously observed in African-derived populations from Central and North America. Genetic variability index (F_{ST} and G_{ST}) were similar to those observed for other African populations analyzed. The overall F_{ST} was 0.073, higher than the 0.042 value estimated by Watkins et al. (2002). However, G_{ST} was 0.067, similar to that observed by Stoneking et al. (1997) i.e. 0.088 in Africans. These were higher than those observed for other world populations, with the exception of Amerindians (Stoneking et al. 1997; Watkins et al. 2002).

Within Amerindians, G_{ST} values obtained with Alu insertions range from 0.102–0.452. Further a comparative study about Asian, Northern Arctic and Amerindian populations have revealed a decreasing trend of heterozygosity and amount

Table 2: Alu diversity for major groups and 31 populations

Population	*Alu Diversity*	
Africans	**0.3487**	(0.3189–0.3785)
Alur	0.3544	
Biaka Pygmy	0.3073	
Hema	0.3503	
!Kung	0.3390	
Mbuti Pygmy (Coriell)	0.3221	
Mbuti Pygmy	0.3135	
Nande	0.3393	
Nguni	0.3445	
Sotho/Tswana	0.3411	
Tsonga	0.3510	
E. Asians	**0.3104**	(0.2729–0.3480)
Cambodian	0.2947	
Chinese	0.3178	
Japanese	0.3064	
Malay	0.3256	
Vietnamese	0.2965	
Europeans	**0.2973**	(0.2616–0.3331)
Finns	0.2927	
French	0.3009	
N. European	0.2964	
Poles	0.2798	
Indians	**0.3159**	(0.2803–0.3514)
Brahmin	0.3128	
Irula	0.3068	
Kapu	0.3117	
Khonda Dora	0.3050	
Kshatriya	0.3031	
Madiga	0.3103	
Mala	0.3113	
Maria Gond	0.3029	
Relli	0.3220	
Santal	0.3007	
Vysya	0.2993	
Yadava	0.3127	

(95% CI in parentheses)

of gene flow was observed from Asian to Amerindian populations (Battilana et al. 2006). Overall, Brazilian populations showed a higher degree of gene flow than average when compared to other worldwide populations, but similar to other African-derived populations.

Alu Polymorphism in Indian Populations

A number of studies have investigated Alu polymorphism among the Indian populations along with other world populations to understand the extent of genetic differentiation and to trace their ancestry. One such study investigated various populations from Africa, East Asia, Europe, and India using 100 Alu markers (Watkins et al. 2003). This study has included eight Indian caste groups and four endogamous south Indian tribal populations. In the NJ network South Indian caste and tribal populations are located between the European and E. Asian clusters. The bootstrap value reflects the fact that the genetic distance between the Indian and E. Asian population groups is almost identical to the genetic distance between the Indian and European population groups. The Indian castes from the state of Andhra Pradesh have exhibited low between-group differences that are probably attributable, in part, to low geographic distances between groups. The tribal Indian groups show relatively high between-group differentiation, which probably can be attributed to reproductive isolation and drift, consistent with previous studies of such populations (Das et al. 1996). Ravindranath et al. (2005) investigated Yadava population from Visakhapatnam, Andhra Pradesh using only four Alu insertion/deletion polymorphisms. They have compared their data with the available data of two Andhra Pradesh tribal populations and concluded that the Yadavas are closer to Konda Reddy than Koya Dora populations.

Vishwanathan et al. (2004) studied genetic structure and affinities of five Dravidian-speaking tribal populations inhabiting the Nilgiri hills of Tamil Nadu, in south India, using 24 autosomal DNA markers, including 7 Alu markers. They have observed that these populations may share a common ancestry and are more closely related to other Indian groups than they are to African groups. Kanthimathi et al (2007) recently studied early immigrant Thevar populations of Tamil Nadu, which are traditionally agriculturists, culturally homogenous and endogamous, using a panel of seven Alu elements. High levels of polymorphism at all loci in their populations reflected their common ancestry. Majumder et al. (1999) studied seven Alu Indel polymorphic markers among 396 unrelated individuals belonging to 14 ethnic populations from different geographical locations and of different socio-cultural hierarchy that included 4 tribes. The Indian populations show not only higher degree of genomic differentiation among but also greater levels of heterozygosity than among most other global populations except Africa.

Population diversity analysis based upon these studies reveal that our perception of the amount and distribution of human genetic diversity is becoming radically altered by the introduction of sophisticated molecular techniques into the field of evolutionary biology. Here Alu insertion polymorphism may play very important role as a genetic marker that has come up with increased precision in estimating the evolutionary relationships between populations.

We have also studied 9 populations from north India using a panel of 10 Alu markers and have seen that the nine north Indian populations are genetically closer to the Caucasian than to Asian populations. The phylogenetic reconstruction of these populations along with other world populations based on only seven Alu markers i.e. ACE, TPA25, PV92, APO, FXIIIB, D1 and B65 has shown that all the nine north Indian populations have clustered with the European populations (Fig. 3). In this phylogenetic analysis a hypothetical population having zero Alu element insertion at each locus was also included and was placed within the cluster of African populations (Fig. 3). The placing of hypothetical population among African populations is further confirmation of African populations being most ancient.

This was further confirmed in a two dimensional principal coordinate (PC) analysis where our populations were compared with other world populations. The first two principal components account for 72.1% of the variance observed (53% and 19.1% respectively) and demonstrated the clustering of the Indian populations between the Asian and European populations (Fig. 4).

Indian sub-continent has witnessed a massive gene flow from varied sources during the history. To predict gene flow we have plotted average heterozygosity of our populations along

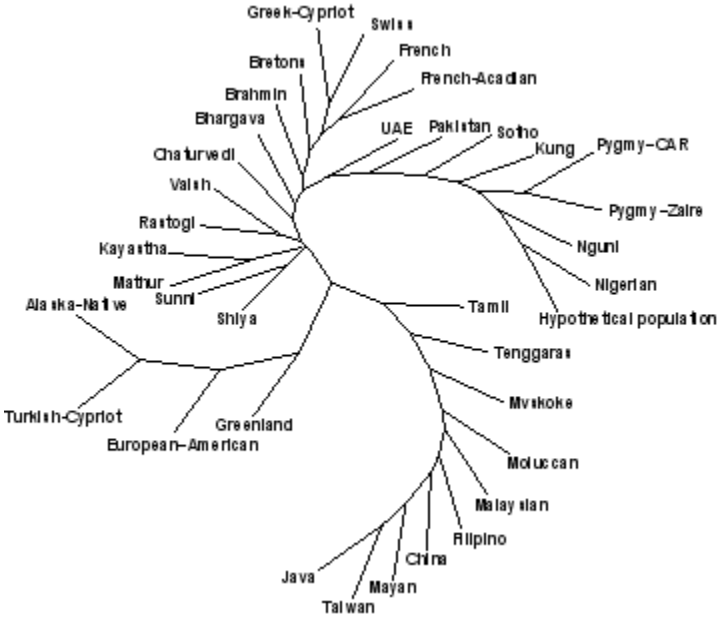

Fig. 3. Phylogram showing Indian populations among world populations

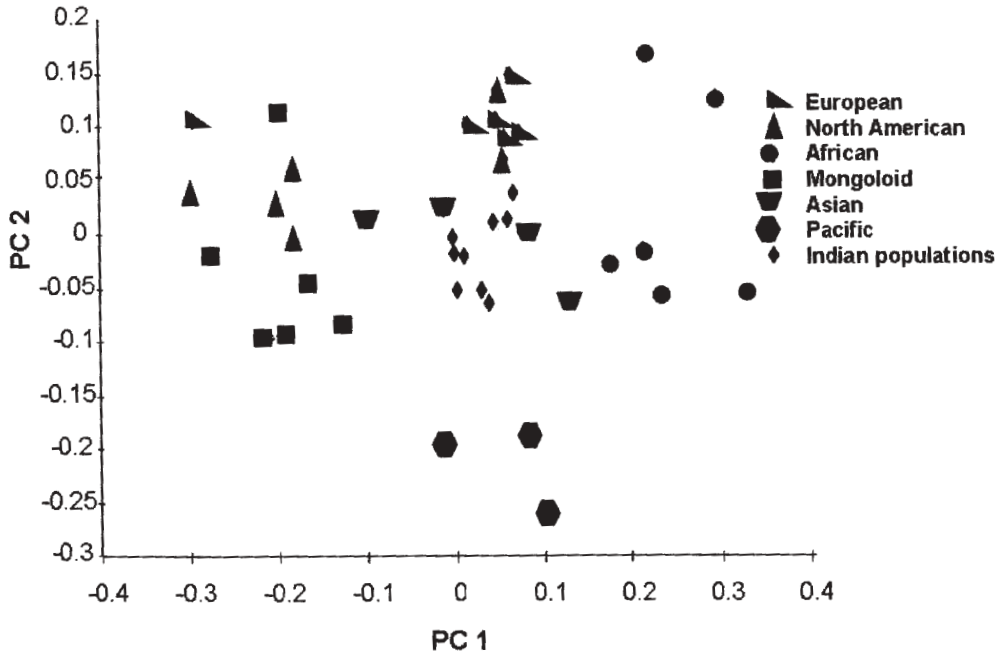

Fig. 4. Principal Component plot of Indian populations with other world populations

with other Indian populations based on seven Alu polymorphic markers against genetic distance from centroid. In this regression plot, all the populations including the three upper Hindu caste populations i.e. Brahmin, and it's two subgroups Bhargava and Chaturvedi were placed marginally above the theoretical regression line, showing relatively higher heterozygosity and thus the higher gene flow than predicted (Fig. 5). Among these populations along with high gene flow a considerable extent of gene differentiation is also observed as suggested by their large distances from the centroid. The explanation could be that gene flow occurred prior to the subdivision of these populations into largely endogamous subpopulation. It is suggested that after migration of modern humans from Africa, there were many rapid population expansions following an initial period of isolation. The studied Indian populations have higher heterozygosity as compared to African populations. This increased heterozygosity in conjunction with greater differentiation could be a result of gene flow with subsequent expansion and subdivisions into smaller endogamous populations leading to rapid genetic differentiation.

The high-resolution analysis of Alu markers of human genome infers that the north Indians are genetically highly diverse people with most of the variation scattered between individuals. The genetic differentiation does exist between endogamous groups across Indian mainland but the differentiation is mainly configured geographically and linguistically, and the stringent socio-cultural norms governing endogamy have only a trivial contribution. A number of migrations and invasions through the northwestern corridor of India have more genetic evidences in the paternal lineages. Despite experiencing a glut of human migrations, the paternal genealogy of north Indians still sustains the records of each of the migratory episodes and the legacy of Indo-Aryan speakers is discernible in the gene pool of each of the studied north Indian populations. Overall, the genetic configuration of Indians is as complex as the history of Indian mainland, interwoven in numerous threads of unknown facts. More genetic data from this part of the world and other critically important regions like Afghanistan, Iran and Iraq are essential in tracing the missing blocks of the causes and consequences of human genetic variation.

To conclude we say that the use of Alu indel polymorphisms is very important tool for population genetic studies because they are identical by descent. Also its use provides direction of mutational change as the ancestral

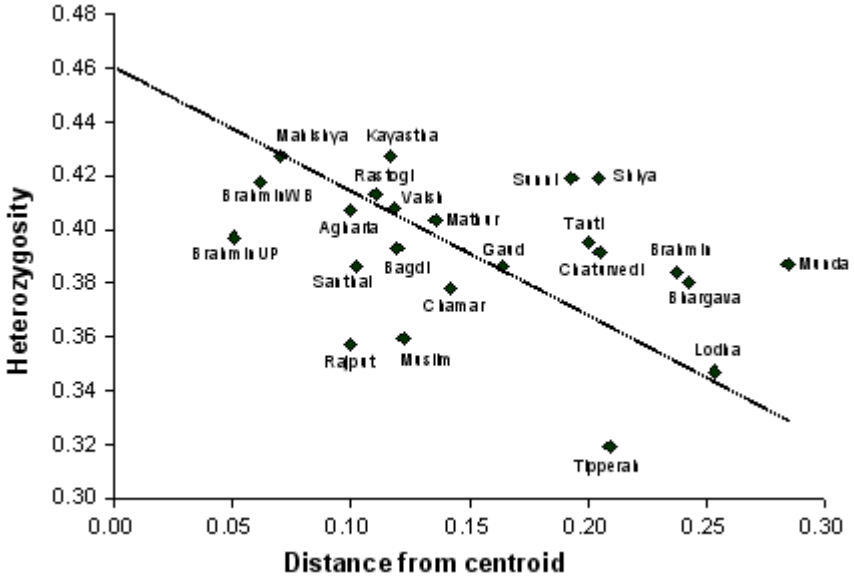

Fig. 5. Plot of genetic distance from centroid vs average heterozygosity

state is the absence of the insertion and newer allele would be the gain of the Alu element at a particular locus. These pieces of information enable us to accurately analyze the population relationships which is generally not possible for other markers. These features make Alu insertion polymorphisms more useful than other autosomal markers for the study of population history and structure.

REFERENCES

Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB et al 2001. Genetic evidence on the origins of Indian caste populations. *Genome Res*, **11:** 994–1004.

Bamshad M, Wooding SP 2003. Signatures of natural selection in the human genome. *Nat Rev Genet*, **4:** 99-111.

Bamshad M, Wooding S, Salisbury BA, Stephens JC 2004. Deconstructing the relationship between genetics and race. *Nat Rev Genet*, **5:** 598-609.

Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA, Jorde LB 2003. Human population genetic structure and inference of group membership. *Am J Hum Genet*, **72:** 578-589.

Battilana J, Fagundes NJ, Heller AH, Goldani A, Freitas LB, Tarazona-Santos E 2006. Alu insertion polymorphisms in Native Americans and related Asian populations. *Ann Hum Biol*, **33:** 142-160.

Batzer MA, Deininger PL 2002. Alu repeats and human genomic diversity. *Nat Rev Genet*, **3:** 370-379.

Batzer MA, Deininger PL, Hellmann-Blumberg U, Jurka J, Labuda D, Rubin CM et al 1996. Standardized nomenclature for Alu repeats. *J Mol Evol*, **42:** 3-6.

Bowcock AM, Kidd JR, Mountain JL, Hebert JM, Carotenuto L, Kidd KK, Cavalli-Sforza LL 1991. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proc Natl Acad Sci USA*, **88:** 3839–3343.

Brookfield JF 2001. Selection on Alu sequences? *Curr Biol*, **11:** R900-901.

Callinan PA, Wang J, Herke SW, Garber RK, Liang P and Batzer MA 2005. Alu Retrotransposition-mediated Deletion. *J Mol Biol*, **348:** 791-800.

Cann RL 2001. Genetic clues to dispersal in human populations: retracing the past from the present. *Science*, **291:** 1742-1748.

Cavalli-Sforza LL, Menozzi P, Piazza A 1994. *The History and Geography of Human Genes.* Princeton: Princeton University Press.

Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bosch E, Martinez-Arias R et al 1998. Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *Am J Hum Genet*, **63:** 1824-1838.

Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M et al 2000. Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet*, **107:** 312-319.

Cordaux RD, Hedges J, Batzer MA 2004. Retrotransposition of Alu elements: how many sources? *Trends in Genetics*, **20:** 464-467.

Cordaux R, Lee J, Dinoso L, Batzer MA 2006. Recently integrated Alu retrotransposons are essentially neutral residents of the human genome. *Gene*, **373:** 138-144.

Cotrim NH, Auricchio MT, Vicente JP, Otto PA, Mingroni-Netto RC 2004. Polymorphic Alu insertions in six Brazilian African-derived populations. *Am J Hum Biol*, **16:** 264-277.

Crawford DC, Schwartz CE, Meadows KL, Newman JL, Taft LF et al 2000. Survey of the fragile X syndrome CGG repeat and the short-tandemrepeat and single-nucleotide-polymorphism haplotypes in an African American population. *Am J Hum Genet*, **66:** 480–493.

Das K, Malhotra KC, Mukherjee BN, Walter H, Majumder PP, Papiha SS 1996. Population structure and genetic differentiation among 16 tribal populations of central India. *Hum Biol*, **68:** 679-705.

Deininger PL, Batzer MA 1999. Alu repeats and human disease. *Mol Genet Metab*, **67:** 183-193.

Deininger PL, Batzer MA 2002. Mammalian retroelements. *Genome Res*, **12:** 1455-1465.

Deininger PL, Daniels G 1986. The recent evolution of mammalian repetitive DNA elements. *Trends Genet*, **2:** 76–80.

Dutta R, Reddy BM, Chattopadhyay P, Kashyap VK, Sun G, Deka R 2002. Patterns of genetic diversity at the nine forensically approved STR loci in the Indian populations. *Hum Biol*, **74:** 33-49.

Feng Q, Moran JV, Kazazian HH, Boeke JD 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, **87:** 905-916.

Fuhrman SA, Deininger PL, LaPorte P, Friedmann T, Geiduschek EP 1981. Analysis of transcription of the human Alu family ubiquitous repeating element by eukaryotic RNA polymerase III. *Nucleic Acids Res*, **9:** 6439-6456.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumensteil B et al 2002. The structure of haplotype blocks in the human genome. *Science*, **296:** 2225-2229.

Gilbert N, Lutz-Prigge S, Moran JV 2002. Genomic deletions created upon LINE-1 retrotransposition. *Cell*, **110:** 315-325.

Goldstein DB, Zhivotovsky LA, Nayar K, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW 1996. Statistical properties of the variation at linked micro-satellite loci: implications for the history of human Y chromosomes. *Mol Biol Evol*, **13:** 1213-1218.

Hamdi H, Nishio H, Zielinski R, Dugaiczyk A 1999. Origin and phylogenetic distribution of Alu DNA repeats: Irreversible events in the evolution of primates. *J Mol Biol*, **289:** 861–871.

Han K, J Xing, Wang H, Hedges DJ, Garber RK, Cordaux R, Batzer MA 2005. Under the genomic radar: The Stealth model of Alu amplification. *Genome Res*, **15:** 655-664.

Heinz CB 1999. *Asian Cultural Traditions. Prospect Heights*. Illinois: Waveland Press, Inc., pp. vvii-364

Hirszfeld L, Herszfeld H 1919. Essal d'application des methods au probleme des races. *Anthropologie*, **29:** 505-517.

Houck CM, Rinehart FP, Schmid CW 1979. A ubiquitous family of repeated DNA sequences in the human genome. *J Mol Biol*, **132:** 289–306.

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.

Ingman M, Kaessmann H, Paabo S, Gyllensten U 2000. Mitochondrial genome variation and the origin of modern humans. *Nature*, **408:** 708–713.

Jin L, Baskett ML, Cavalli-Sforza LL, Zhivotovsky LA, Feldman MW, Rosenberg NA 2000. Microsatellite evolution in modern humans: a comparison of two data sets from the same populations. *Ann Hum Genet*, **64:** 117-134.

Jorde LB, Rogers AR, Bamshad M, Watkins WS, Krakowiak P, Sung S et al 1997. Microsatellite diversity and the demographic history of modern humans. *Proc Natl Acad Sci USA*, **94:** 3100-3103.

Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA 2000. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet*, **66:** 979-988.

Kanthimathi S, Vijaya M, Srikumari CR, Govinda Reddy P, Majumder PP, Ramesh A 2007. A genetic structure of the early immigrants (Mukkalathor) of Tamil Nadu as inferred from autosomal loci. *Int J Hum Genet*, **7:** 167-173.

Lahr MM, Foley RA 1998. Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. *Am J Phys Anthropol Suppl*, **27:** 137-176.

Lakshmi N, Demarchi DA, Veerraju P, Rao TV. 2002. Population structure and genetic differentiation among the substructured Vysya caste population in comparison to the other populations of Andhra Pradesh. India. *Ann Hum Biol*, **29:** 538-549.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W et al 2001. Initial sequencing and analysis of the human genome. *Nature*, **409:** 860-921.

Luan DD, Korman MH, Jakubczak JL, Eickbush TH 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non- LTR retrotransposition. *Cell*, **72:** 595-605.

Majumder PP, Roy B, Banerjee S, Chakraborty M, Dey B, Mukherjee N et al 1999. Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. *Eur J Hum Genet*, **7:** 435–446.

McClintock B 1956. Controlling elements and the gene. *Cold Spring Harb Symp Quant Biol*, **21:** 197-216.

Mills RE, Bennett EA, Iskow RC, Devine SE 2007. Which transposable elements are active in the human genome? *Trends Genet*, **23:** 183-191.

Misra VN 2001. Prehistoric human colonization of India. *J Biosci*, **26:** 491–531.

Mizuuchi K 1992. Transpositional recombination: mechanistic insights from studies of mu and other elements. *Annu Rev Biochem*, **61:** 1011-1051.

Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH 1996. High frequency retrotransposition in cultured mammalian cells. *Cell*, **87:** 917-927.

Nasidze I, Risch GM, Robichaux M, Sherry ST, Batzer MA, Stoneking M 2001. Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus. *Eur J Hum Genet*, **9:** 267-272.

Okada N. 1991. SINEs. *Curr Opin Genet Dev*, **1:** 498-504.

Ovchinnikov IV, Gotherstrom A, Romanova GP, Kharitonov VM, Liden K, Goodwin W 2000. Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature*, **404:** 490-493.

Papiha SS 1996. Genetic variation in India. *Hum Biol*, **68:** 607-628.

Pauling L, Itano HA, et al 1949. Sickle cell anemia a molecular disease. *Science*, **110:** 543-548.

Quintana-Murci L, Krausz C, Zerjal T, Sayar SH, Hammer MF, Mehdi SQ et al 2001. Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am J Hum Genet*, **68:** 537–542.

Ravindranath VM, Lakshmi N, Ramesh M, Veerraju P 2005. Alu Insertion / Deletion Polymorphisms in Yadava Population of Andhra Pradesh, South India. *Int J Hum Genet*, **5:** 223-224.

Relethford JH 2001. Ancient DNA and the origin of modern humans. *Proc Natl Acad Sci USA*, **98:** 390-391.

Renfrew C 1989. The origins of Indo-European languages. *Sci Am*, **261:** 82-90.

Rieder MJ, Taylor SL, Clark AG, Nickerson DA 1999. Sequence variation in the human angiotensin converting enzyme. *Nat Genet*, **22:** 59-62.

Rogers J 1998. Retroposons defined. *Nature*, **301:** 460.

Romualdi C, Balding D, Nasidze IS, Risch G, Robichaux M, Sherry ST, Stoneking M, Batzer MA, Barbujani G 2002. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res*, **12:** 602–612.

Rosenberg NA, Burke T, Elo K, Feldman MW, Freidlin PJ, Groenen MA et al 2001 Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics*, **159:** 699-713.

Rousset F 1996. Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics*, **142:** 1357-1362.

Roy-Engel AM, Carroll ML, Vogel E, Garber RK, Nguyen SV, Salem AH, Batzer MA, Deininger PL 2001. Alu insertion polymorphisms for the study of human genomic diversity. *Genetics*, **159:** 279-290.

Ruhlen M 1991. A guide to the world's languages. In: A Sherratt (Ed.). *The Cambridge encyclopedia of archaeology.* London: Edward Arnold.

Sen SK, Han K, Wang J, Lee J, Wang H, Callinan PA et al 2006. Human genomic deletions mediated by recombination between Alu elements. *Am J Hum Genet*, **79:** 41-53.

Shriver MD, Kennedy GC, Parra EJ, Lawson HA, Sonpar V, Huang J et al 2004. The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Hum Genomics*, **1:** 274-286.

Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot

SS, Saha N, et al 1997. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res*, **7:** 1061–1070.

Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G, Boeke JD 2002. Human l1 retrotransposition is associated with genetic instability in vivo. *Cell*, **110:** 327-338.

Ullu E, Tschudi C 1984. Alu sequences are processed 7SL RNA genes. *Nature*, **312:** 171-172.

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH et al 2000. Y chromosome sequence variation and the history of human populations. *Nat Genet*, **26:** 358-361.

Vishwanathan H, Deepa E, Cordaux R, Stoneking M, Usha Rani MV, Majumder PP 2004. Genetic structure and affinities among tribal populations of southern India: A study of 24 autosomal DNA markers. *Annals of Human Genetics*, **68:** 128-138.

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P et al 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420:** 520-562.

Watkins WS, Ostler CT, Brassington A, Bamshad MJ, Batzer MA, Carrol ML, Nguyen SV, Jorde LB 2002. Population structure and distribution of 100 polymorphic Alu insertion polymorphisms. *Am J Hum Genet*, **71:** 354.

Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AM et al 2003. Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. *Genome Res*, **13:** 1607-1618.

Weber JL, Wong C 1993. Mutation of human short tandem repeats. *Hum Mol Genet*, **2:** 1123-1128.

Wilson JF, Weale ME, Smith AC, Gratrix F, Fletcher B, Thomas MG et al 2001. Population genetic structure of variable drug response. *Nat Genet*, **29:** 265–269.

Wolpoff MH, Hawks J, Frayer DW, Hunley K 2001. Modern human ancestry at the peripheries: a test of the replacement theory. *Science*, **291:** 293-297.

Yu N, Zhao Z, Fu Y-X, Sambuughin N, Ramsay M, Jenkins T et al 2001. Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. *Mol Biol Evol*, **18:** 214-222.

Zhivotovsky LA, Goldstein DB, Feldman MW 2001. Genetic sampling error of distance (delta(mu))2 and variation in mutation rate among microsatellite loci. *Mol Biol Evol*, **18:** 2141-2145.