Deposited research article

# Migration events play significant role in genetic differentiation: A microsatellite-based study on Sikkim settlers

Saurav Guha[a], R.Trivedi[a] and V.K.Kashyap[a, b*]

Addresses: [a] Central Forensic Science Laboratory, Kolkata, India. [b] National Institute of Biologicals, Noida, India.

Correspondence: V.K. Kashyap. E-mail: vkk2k@hotmail.com and sauravguhain@yahoo.com

Posted: 3 June 2005

*Genome* ***Biology*** 2005, **6**:P9

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2005/6/7/P9



Received: 31 May 2005

This is the first version of this article to be made available publicly and no other version is available at present.

# Migration events play significant role in genetic differentiation: A microsatellite-based study on Sikkim settlers.

Saurav Guha[a], R.Trivedi[a] and V.K.Kashyap[a, b] *

[a] Central Forensic Science Laboratory, Kolkata

[b] National Institute of Biologicals, Noida

Key Words : Microsatellite, Migration, Sikkim, Mongoloids, Genetic Diversity, Genetic Distance.

Running title: Microsatellite Polymorphism of Sikkim Settelers.

*Corresponding author :-
National Institute of Biologicals,

A 32, Sector 62 (Institutional Area)

Noida - 201307, India

TEL:- +91-120-2400027, FAX:- +91-120-2400027,

E-MAIL: vkk2k@hotmail.com and sauravguhain@yahoo.com

**Abstract**

**Background:**

A wide spectrum of genetic diversity in mongoloids of India is well documented. Though all mongoloids of India are known to have originated from the Mongol region of China but the period and route of migration from their native land to different Himalayan regions is little known. Thus the studies on genomic diversity of people of Sikkim, a central Himalayan state of India with different migrant mongoloid groups, assume great significance in understanding the impact of migratory events in the genetic differentiation of populations. We therefore studied the genetic diversity on the basis of 13-tetra nucleotide and 2 penta nucleotide microsatellite loci for a total of 208 allele frequencies in three major populations of Sikkim, with different ethno history and time of settlement.

**Result:**

The study on microsatellite allele frequency data suggests that all the three populations of Sikkim are genetically more akin to the mongoloids of China and distinctly apart from the mongoloids of Northeast India. However Sikkim populations are also genetically close to non-mongoloids of surrounding areas. The average heterozygosity and coefficient of gene differentiation among Sikkim populations are moderate. Number of shared alleles and their frequencies, time of divergence and bottleneck effect reveal a distinctiveness of the mongoloids settled in Sikkim from the main Indian mongoloid stock as also different route of migration than the mongoloid population of Northeast India.

**Conclusion:**

Our study clearly demonstrates that the present day mongoloids of Sikkim are genetically distinct from mongoloids of Northeast India due to their different route of migration, time of settlement, and admixture with other non-mongoloid populations of adjoining areas. This substantiates that migratory events have played a significant role in the differentiation of mongoloids of India.

**Background**

The origin, dispersal and antiquity of Homo sapiens in the Asian peninsula have attained high significance after the discovery of new archaeological and anthropological evidences from East Asia. The extensive genetic information present on mongoloids, a major human ethnic group of eastern and southeastern Asia, also provides a deep insight into the origin of modern human and the time of their dispersal in different continents.

Mongoloids constitute about one-fourth of the world population and exhibit a high level of diversity. In India, mongoloids contribute approximately 3% of the total population, mainly inhabiting the hills and adjoining plains of Northeast and Central Himalayas. Mongoloids in India are originally the migrant groups and their settlements in different regions of the country were not as a single influx; rather, a process of larger or smaller waves of migration. These migrant populations interacted with the non-mongoloid populations due to geographical proximity. Because of different periods of origin from various regions, migration and settlement, mongoloids of India differentiated in different socio-linguistic cultural groups which have impacted their genetic structure.

Among the different areas of mongoloid ethnicity in India, Sikkim – an Indian state which is geographically a buffer zone between mongoloid and non-mongoloid populations of South Asia -- is unique in its population structure. Geographically Sikkim is surrounded by the Royal Kingdom of Nepal in the west, the Royal Kingdom of Bhutan in the east, China in the north with mongoloid populations and the Indian state of West Bengal in the south with both mongoloid non-mongoloid populations. Moreover, the genetic structure of Sikkim population with their biological affinities, origins and dispersal is important in understanding the impact of migration on the genetic structure of Mongoloids in India. The Nepali, Bhutia and Lepcha -- the three major populations in this Himalayan land -- represent the Indo-

mongoloid ethnic group, albeit the Lepcha are said to be the forerunner among the settlers in the state, followed by Bhutia and Nepali migration in the state [1]. The Bhutia community is basically the people of Tibetan (Chinese) origin, and their migration to Sikkim had occurred at least 800 years ago. The Nepalese arrived in Sikkim about 200 years ago [1]. Though the migration of Lepcha to Sikkim is a contentious issue, there are two major theories about the Lepcha origin and period of migration to Sikkim: (i) A small population of Naga stock from the Garo hills of Assam migrated to Sikkim about 2000 years ago, and (ii) a group from east Asia mainland migrated through Tibet (China) during an earlier phase of human migration to the Himalayas at least 3000 years ago [1].

The complex migration and settlement history of different groups in Sikkim with respect to their genetic history is still little known. Autosomal and Y chromosome STR markers [2-4], as well as Mitochondrial DNA [5] studies illustrate migration and genetic diversity of mongoloids in East and South-East Asia. These studies, however, do not provide any information about mongoloid migration and settlement in different regions of India. Although, several studies have been carried out on genetic diversity, phylogenetic relationship, and the pattern of gene flow among Indian mongoloid populations based on classical genetic markers, e.g. Tf, Ge, PGM1 loci [6, 7], serum proteins and red cell enzymes [8], and a combined study by Roychoudhury and Nei. [9], they fail to provide any insight into the history of migration and settlement of Mongoloids in India. Similarly the study with 17 polymorphic system of the blood by Bhasin et al. 1986 provides information about the genetic structure of people of Sikkim to some extent.

In this study we used microsatellite DNA markers to address the consequence of migratory events on genetic structure of Sikkim populations as, they are more abundant in the

genome vis-à-vis classical genetic markers [10], have high heterozygosity [11] and polymorphism is ubiquitous even in inbred populatiosn or species [12]. Microsatellites are easily amenable to automated procedure of typing [13] and to statistical tools used for other markers [14, 15]. The analytical techniques of quantitative genetics are being applied to microsatellite alleles, as they are characterized quantitatively by their size, i.e. number of repeat of DNA motif. Measures of population subdivision [16] and distance between the populations [17, 18] as well as higher statistics [19, 20] have been extensively employed for the study of microsatellite polymorphisms.

The above advantages have made microsatellite markers extremely informative in understanding the genetic structure, migratory history and evolution of human populations [21 - 23] and our marker of choice to examine the impact of migratory events on the genetic differentiation of Sikkim settlers and their genetic relationship with other mongoloids of Northeast India, and Caucasoid-affiliated populations of adjoining area.

## Results

*Genetic differentiation & heterozygosity: -*

Locus and population wise heterozygosity and Gst (coefficient of gene differentiation) values reflecting the extent of differentiation among the populations of Sikkim are shown in Table 1. The average Gst value (0.0237) suggests moderate degree of gene differentiation among the studied populations. The values however vastly differ from loci to loci; it is only 0.004 at loci D3S1358, and D16S539, while 0.088 at loci FGA. Similarly moderate differentiation (0.0206 to 0.0365) was observed at loci Penta D, CSF1PO, TPOX, D8S1179, VWA, D5S818 and D7S820. Average heterozygosity observed at different loci depicts the

extent of variation among the Sikkim populations (Table 1). The average heterozygosities of three populations were 0.786 (Nepali), 0.747 (Bhutia) and 0.684 (Lepcha). Among the 15 loci, highest heterozygosity was observed at locus Penta E (0.841-0.902) and lowest at locus TPOX (0.546-0.634).

*Allele distribution & Variance: -*

In general, the number of shared alleles and their frequencies among 12 populations vastly differ from loci to loci, (Table 2) (no allele frequencies are given). Of the twelve microsatellite loci (excluded PentaD, PentaE and D16S539), the number of alleles present at loci D21S11 (V= 2.239), vWA (V= 2.78), D8S1179 (V=3.77) D18S51 (V=6.38) and FGA (V= 3.98) are quite varied among the populations whereas other loci demonstrate a narrow range of variation (V <2.000). In D21S11 and FGA the number of shared alleles ranges from (6- 16) to (7-18) respectively in different populations. The Garo exhibit only 6 alleles (V= 0.892) while the Lepcha have 16 alleles (V= 2.93) at D21S11. The Lepchas show only 7 alleles (V= 1.123) at FGA locus at which the Nepali exhibit 18 alleles (V= 5.096). At D18S51 the Naga exhibit 10 alleles (V= 5.892) whereas Nepalese exhibit a maximum of 15 alleles,(V= 8.736). The variations of numbers of shared alleles are relatively stable for different loci in non-mongoloids and Chinese compared to Indo-mongoloids. The frequencies of shared alleles fluctuate from population to population. At the tetra nucleotide repeat locus THO1, the allele frequency distribution is quite varied among populations. Repeat 9 is the predominant allele in every Asian mongoloid population with frequency of 0.40 to 0.52 [24], which is also observed in all the three populations of Sikkim viz. the Nepali (0.492), Bhutia (0.343) and Lepcha (0.500). But repeat 9.3 not 9 shows a noticeable frequency ($\geq$ 0.300) among the mongoloid population of Northeast India, which is observed in much lower

frequency (<0.15) in Sikkim populations like other Asian mongoloid populations [24]. At locus D13S317 the allele 18 is also present with moderate frequency (>0.200) in mongoloid populations of Northeast India compared to other populations.

*Bottleneck test: -*

We performed SIGN test for three Sikkim populations to find out the extent of bottleneck effect that these populations might have experienced. In a recently bottlenecked population the observed gene diversity (He) is higher than expected equilibrium gene diversity (Heq) if the loci are evolving under the IAM [25]. If the loci evolve under the SMM, then there can be a situation when heterozygosity excess (He>Heq) is not being observed [26]. The SIGN test results under IAM and SMM assumptions are presented in Table 3. In Nepali population under IAM 14 loci showed heterozygosity excess (He>Heq) with (P= 0.00484) and under SMM, 8 loci showed (He>Heq) with (P= 0.41469). Bhutia population exhibited 13 loci with (He>Heq) (P=0.02380) under IAM and in SMM 5 loci with (He>Heq), (P=0.04238). Under IAM Lepcha showed 9 loci with (He>Heq) (P=0.57428) and in SMM 2 loci with (He>Heq) (P=0.00037). These values indicate a significant level of bottleneck within these populations during their period of migration to different regions.

*Genetic Distance & Phylogeny: -*

The genetic distances among the populations were computed by employing different genetic distance measures viz. $D_A$, Ds, Dc, Dsw, Fst and $(\delta\mu)^2$ for allele frequencies of 12 STR loci. The NJ, (Figure 1) and UPGMA (not shown) phylogenetic construction with these distances were quite similar in which Naga, Kuki, Garo and Hmar form a different cluster with 96-97% bootstrap value separated from the main cluster. In NJ tree with distances $D_A$, Ds, Dc, Dsw, and Fst, the Nepali, Bhutia and Lepcha form a cluster with Chinese and

Caucasoid-- affiliated Brahmins of India, whereas with $(\delta\mu)^2$ Garo population form a cluster with Caucasoid. In this study we present NJ tree for Dc and Da since they are more precise for obtaining the correct tree topology [27].

*Time of Divergence: -*

The time of divergence of Lepcha from Naga was calculated by using Ds, Dsw and $(\delta\mu)^2$ distances as these distances are linear to evolutionary time [27] and average microsatellite mutation rate i.e. $5.6 \times 10^{-4}$ per generation [28].The separation time for these populations is varied for different distance calculations, it is 1718 years for Dsw, 3730 years for Ds and 8303 years for $(\delta\mu)^2$ . For our study we find Ds to be more suitable for divergence time calculation since $(\delta\mu)^2$ and Dsw are not precise if the population size is not constant throughout evolution and has experienced bottleneck. It also exhibits high heterozygosity level (nearly 0.800) with small divergence level among the populations that are connected by weak gene flow [27, 29].

**Discussion**

The migration of mongoloids from their native land is known to occur in different phases, periods and directions. Reportedly, at least 6.5 Ky bp years ago a small group of population from the Sino-Tibetan linguistic family moved from the valley of Yellow river towards west, and then turned to south to southwestern direction [30]. Tibeto-Burman speakers, a sub-family of Sino-Tibetan linguistic family, followed two main routes of dispersal: (i) towards west in Tibet (China) and then down to Nepal, Sikkim, Bhutan and Northern India, and (ii) towards southwest down the river valleys along the eastern edge of the Tibetan plateau through the 'Ethnic Corridor'. During migration there have been several

contacts between northern and central Asian languages with Tibeto-Burman languages [31]. These language contacts were responsible for genetic affinity of Tibeto-Burman linguistic families with non Tibeto-Burman linguistic families of adjoining areas.

Mongoloids of Central Himalayas and Northeast India belong to two different linguistic subfamilies under Tibeto-Burman sub linguistic family. Sikkim populations belong to Himalayish linguistic group (Indo-sphere) while the northeastern Indian Mongoloid populations belong to Jingpho-konyak-bodo and kuku-chin-naga group (Sino-sphere) under Tibeto-Burman linguistic subfamily [32]. These sub-language groups have been developed as the result of a close contact of mongoloids with surrounding non-mongoloids. The Indo-sphere group has a strong influence from Indic and Dravidian languages of India [32].

Phylogenetic trees based upon 12 microsatellite markers clearly depict a pattern of co-evolution of genes and languages in the studied population. The clustering pattern of NJ tree with Da, Ds, Dc, Dsw and Fst distances explicitly demonstrated, Sikkim populations' genetic affinity to Caucasoid--affiliated Brahmins of India rather than mongoloids of Northeast India.. The Sikkim populations exhibit a clear clustering pattern with Chinese population as both Sikkim and Chinese populations have experienced gene flow from Caucasian population during their migration from the mainland of east Asia through different routes. Mongoloids of Northeast India (Naga, Kuki, Hmar and Garo) constitute a separate cluster with 96-97% bootstrap value and are phylogenetically distant from Mongoloids of Sikkim, and China as well as from Caucasoid affiliated Brahmin population of India.

The numbers of shared alleles distribution and their variance within loci fluctuate from population to population, which is highly noticeable at loci D21S11, D18S51, D8S1179, VWA and FGA. This distribution pattern clearly depicts a large number of variations present

in Indo-mongoloid populations whereas non-mongoloids and Chinese populations are quite stable. Number of shared allele variation occurs either due to genetic drift in which it decreases or gene flow, which introduces new alleles and causes the variation to rise. These stochastic events basically affect the loci with large number of alleles and as expected FGA (number of alleles 28), D21S11 (25), D18S51 (23), D8S1179 (12) and VWA (13) loci showed highest range of variation. In mongoloid populations of Northeast India, the presence of allele 9.3 at THO1 locus in higher frequency in comparison to other Asian mongoloid populations indicates a genetic drift experienced by these populations during migration, which perhaps increased its allele frequency.

In bottleneck test we used both the mutation models, IAM and SMM, with the assumption that the population maintains mutation-drift equilibrium for these markers. In IAM model Nepali population exhibited 14 loci with heterozygosity excess (He>Heq) and statistical significant probability value whereas in SMM 8 loci showed heterozygosity excess (He>Heq) with statistically insignificant probability value. The values suggest that the Nepali population had experienced a bottleneck but rapid population expansion caused exclusive heterozygosity deficiency, which affected the SMM calculation. In Bhutia population both the models disapprove the hypothesis of mutation–drift equilibrium in the population, whereas the Lepcha population depicts a statistically significant value in favor of bottleneck under SMM. However, 13 loci with heterozygosity deficiency indicate presence of large number of alleles in some loci. This may arise due to gap filling mechanism within some loci under SMM, which has been affected due to bottleneck.

In Sikkim populations the lower frequency of 9.3 allele at locus THO1, number of shared alleles and probability values of SIGN test suggest a genetic drift and occurrence of

bottleneck during migration in small groups of mongoloids to Sikkim in early phase of their settlement. The time of divergence (~3730 years) of Lepcha from Naga clan suggests that the Lepcha population was separated from Naga populations in early phase of migration of humans to Himalayas at least 4000-5000 years ago [33]. This suggests Lepcha population does not belong to Naga stock of Northeast India and the route of migration was probably not similar for the Lepchas of Sikkim and Mongoloid populations of Northeast India.

The moderate value of average Gst (0.0237) shows relatively high differentiation among the populations of Sikkim. The average Gst value has been found higher than the observed average Gst value with traditional markers for other Indian populations, which is only 0.015 (both castes and tribes). The moderate value of average Gst is may be due to the high level of endogamy in Lepcha and Bhutia.

On the basis of microsatellite analysis it may be concluded that the present genetic structure of mongoloids of Sikkim is not akin to the mongoloids of Northeast India in spite of their similar ethnicity, and similar habitate. Due to long time isolation and amalgamation with non-mongoloids of adjacent area these communities exhibit a mixed gene pool. The time of divergence reveal the route of the possible migration of Lepcha of Sikkim in India was from East Asia through Tibet. This could be the reason of their high genetic affinity with Nepali, Bhutia, and those populations (e.g. Chinese) who had also migrated via the same route i.e. Zang (Tibet)-main corridor, the most frequent route to enter the Himalayas from the east (Fig. 2).

The extent of genetic variation in three predominant mongoloid populations of Sikkim and mongoloids of northeast India could be attributed to significant role played by migratory events in genetic differentiation. Although a larger number of loci are essential for drawing

the inference on the short term evolution of such populations, our results clearly demonstrate that these STR loci offer new hope of understanding of recent history more precisely than traditional genetic markers and other biological variables.

## Materials and Methods

*Populations and collection of biological specimen:* **-**

The population studied from Sikkim comprised of the Nepali (n = 110), Bhutia (n = 75) and Lepcha (n = 48). These populations belong to the Tibeto-Burman linguistic subfamily and represent the Indo-Mongoloid ethnic group [34]. Blood or buccal swab samples of individuals were randomly collected from eastern district of Sikkim with the full consent of the participants. The allele frequencies of 208 alleles of fifteen microsatellite loci have been published [35].

*Typing of STR Loci: -*

Genomic DNA was extracted by using standard phenol / chloroform procedure [36]. Quantitation of DNA was carried out using the Quantiblot kit (PE Applied Biosystems) and subsequent PCR amplification was performed using the Powerplex$^{TM}$ 16 multiplex System (Promega Corp, Madison, USA). The products were detected on a 5% denaturing polyacrylamide sequencing gels using the ABI Prism$^{TM}$ 377 DNA Sequencer (PE Applied Biosystems) and genotype classification was made by comparison with allelic ladders provided with the Powerplex$^{TM}$ 16 System.

*Reference data: -*

The comparison was done with STR data of various mongoloid populations of Northeast India viz Naga (n = 78), Hmar (n=60), Kuki (n = 75) and Garo (n = 80), [37, 38]

which belong to Tibeto-Burman linguistic subfamily and represent the Indo-Mongoloid ethnic group of Northeast India [34]. Brahmins of Orissa (n = 60) [39], Bihar (n = 59) [40] and Karnataka (n = 65) [41] have been taken to be of the Indo-Caucasian origin. Chinese and Caucasian data have been used for global population reference [42].

*Statistical Analysis: -*

The frequencies of each allele for individual STR loci were calculated from the number of each genotype in the sample set by the method of gene count [43]. We have used 165 alleles of 12 microsatellite loci for comparison, since data on D16S539, Penta E, and Penta D loci are not available for Northeast Indian populations. The number of shared alleles and their variance (V) were scored for the 12 loci in MICROSAT to check the allelic diversity among the populations. Pair-wise genetic distances among populations were computed using various distance measures $D_A$ [44], Ds [14], Dc [45], Dsw [46], $(\delta\mu)^2$ [17] and Fst [47] with 1000 replications using simulation program MICROSAT [48]. These genetic distance measures are seen to be a highly efficient parameter for computing correct phylogeny trees and evolutionary time span under different evolutionary conditions. Besides they are least affected by small sample size [27]. Phylogeny analysis was carried out following the neighbour joining (NJ) [49] and the unweighted pair group method with arithmetic mean (UPGMA) [50] methods. The Gst (gene diversity) and heterozygosity values were computed using the programme DISPAN [51] and programme [52] was used for generating phylogenetic trees. Time of separation from common ancestor has been calculated by using the genetic distances Ds, Dsw and $(\delta\mu)^2$ between populations. SIGN test for 15 microsatellite loci of Sikkim populations was performed to check whether populations experienced any bottleneck in recent past or not [26] under IAM (Infinite allele model) as well as SMM

(Stepwise mutation model) conditions by using the computer simulation programme BOTTLENECK [53] with 1000 replications.

**Acknowledgement**

This research work was supported by a grant under the IX th plan to Central Forensic Science Laboratory, Kolkata, Ministry of Home Affairs, Govt. of India. Saurav Guha is also thankful to BPR & D for fellowship. The expert editorial assistance of Mrs. Seema Singh is acknowledged. The authors also wish to thank all the blood and buccal swab donors who made this study possible.

**References**


1. National Informatics Centre, Sikkim (http://www.sikkim.nic.in)
2. Su B, Xiao C, Deka R, Seieslstad MT, Kangwanpong D, Xiao J, Lu D, Underhill P, Cavalli-Sforza LL, Chakraborty R, Jin L: **Y-chromosome haplotypes reveal prehistorical migrations to the Himalayas**. *Hum Genet* 2000, **107**: 582-590.
3. Li J, Bing Su: **Natives or immigrants: Modern human origin in East Asia**. *Nature Reviews Genetics*, 2000, **1:** 126-133.
4. Bianchi NO, Cafanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, Camelo JSL: **Characterization of ancestral and derived Y-chromosomal haplotypes of new world native populations.** *Am J Hum Genet*. 1998, **63**: 1862-1871.
5. Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen KK, Wallace DC: **Southeast Asia mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations.** *Genetics* 1992, **130:** 139-152
6. Saha N, Bhattacaryya SP, Mukhopadhyay B, Bhattacharyya SK, Gupta R, Basu A: **A genetic study among the Lepchas of the Darjeeling area of eastern India**, *Human Heredity* 1987a, **37**: 113-121.
7. Saha N, Mukhopadhyay B, Bhattacharyya SK, Gupta R, Basu A: **The distribution of transferring, group-specific component and phosphoglucomutase-1 subtypes among the Lepchas of Darjeeling, Eastern India,** *Japanese Journal Of Human genetics* 1987b, **32**: 311-318.
8. Saha N, Hong SH, Wong HA, Tay JSH: **Red cell glucose-6-phosphate dehydrogenase phenotypes in several mongoloid population of eastern India:**

**existence of a non dominant fast variant in two Australian tribes.** *Annals of Human Biology* 1990, **17**: 529-532.

9. Roychoudhury AK, Nei M: **The emergence and dispersal of mongoloids**. *J. Indian Anthrop. Soc*. 1997, **32**:1-49.

10. Gyapay G, Morissette J, Vignal A, Dib C, Fizames C: The 1993-1994 **Genethon human genetic linkage map**. *Nature Genet*. 1994, **7**: 246-339.
11. Weissenbach J, Gyapay G, Dib C, Vignal A, Morrissette P: **A second generation linkage map of the human genome**. *Nature* 1992, **359**:794-801.
12. Gilbert DA, Lehman N, O'brien SJ, Wayne RK: **Genetic fingerprinting reflects population differentiation in California Channel Island fox**. *Nature* 1990, **344**: 764-767.
13. Lin Z, Cui X, Li H: **Multiplex genotype determination at a large number of gene loci.** *Proc. Natl. Acad. Sci. USA* 1996, **93**: 2582-2587.
14. Nei M: *Molecular evolutionary genetics*. Columbia University Press, New York, N.Y. 1987.
15. Weir B: *Genetic data analysis*, Sinauer, Sunderland, Mass. 1990.
16. Slatkin M: **A measure of population subdivision based on microsatellite allele frequencies**. *Genetics* 1995, **139**: 457-462.
17. Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW: **An evaluation of genetic distances for use with microsatellite loci.** *Genetics* 1995a, **139**: 463-471.

18. Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW: **Genetic absolute dating based on microsatellites and the origin of modern human**. *Proc. Natl. Acad. Sci. USA* 1995b, **92**: 6723-6727.

19. Zhivotovsky LA, Feldman MW: **Microsatellite variability and genetic distance**. *Proc. Natl. Acad. Sci. USA* 1995, **92**: 11549-11552.

20. Goldstein DB, Zhivotovsky LA, Nayar K, Linares AR, Cavalli-Sforza LL, Feldman MW: **Statistical properties of the variation at linked microsatellite loci: implication of human Y-chromosomes.** *Mol. Biol. Evol*. 1996, **13**: 1213-1218.

21. Charkaborty R, Jin L: **A unified approach to study hypervariable polymorphisms: statistical consideration of determining relatedness and population distance.** *DNA Fingerprinting State of the Science.* Edited by Pena SDJ, Chakraborty R, Epplen JT, Jefereys AJ: Birakhauser, Basel. 1993, 153-175.

22. Bowcock AM, Linares RA, Tomforhrde J, Minch E, Kidd JR, Cavalli-Sforza LL: **High resolution of human evolutionary trees with polymorphic microsatellites**. *Nature* 1994, **368**: 455-457.

23. Deka R, Shriver MD, Yu LM, Ferrel RE, Chakraborty R: **Intra and inter population diversity at short tandem repeat loci in diverse population of the world**. *Electrophoresis* 1995, **16:**1659-1664.

24. Deka R, Shriver MD: **Genetic variation at twenty three microsatellite loci in sixteen human populations.** *J. Genet*. 1999, **78**: 99-121.

25. Maruyama T, Fuerst PA: **Population Bottlenecks and non equilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck**. *Genetics* 1985, **111**: 675-689.

26. Cornuet JM, Luikart G: **Description and power of analysis of two tests for detecting recent population bottlenecks from allele frequency data**. *Genetics* 1996, **144**: 2001-2014.

27. Takezaki N, Nei M: **Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA**. *Genetics* 1996, **144**: 389-399.

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

29. Zhivotovsky LA: **Estimating Divergence Time with the Use of Microsatellite Genetic Distances: Impacts of Population growth and gene Flow**. *Mol. Biol. Evol*. 2001, **18**: 700-709.

30. Treistman JM: *The prehistory of China. USA*: Natural History Press, 1972.

31. Snellgrove D, Richardson H: *A cultural history of Tibet*. Boston and London, Shambhala, 1986.

32. Lapolla JR: *The roll of migration and language contact in the development of Sino-Tibetan language family. Areal Diffusion and Genetic Inheritance; Case Studies in Language Change.* Oxford: Oxford University Press, 1999.

33. Cavalli-Sforza LL, Piazza MP: *The history and geography of human genes*. Princeton University Press, Princeton, NJ. 1994.

34. Singh KS: *India's Communities, National Series. People of India*. Oxford University Press, 1998.

35. Kashyap VK, Guha S, Trivedi R: **Concordance study on 15 STR loci in Three Major Population of Himalayan State Sikkim**. *J. Forensic Sci.* 2002, **47**:1163-1167.

36. Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning. A Laboratory Manual. 2nd ed.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

37. Dutta R, Kashyap VK: **Genetic variation observed at three tetrameric short tandem repeat loci HUMTHO1, TPOX, and CSF1PO in five ethnic population groups of Northeastern India.** *Am. J. Hum. Biol*. 2001, **13**: 23-29.

38. Chattopadhyay P, Dutta R, Kashyap VK: **Genetic variation observed at Nine Fluorescent labeled STR loci among Four Tribal population group of the Indian sub continent**. *J. Forensic Sci.* 2001, **46:** 201-205.

39. Sahoo S, Kashyap VK: **Allele frequency of data for Powerplex 16 loci in four major population of Orissa, India**. *J. Forensic Sci.* 2002, **47**: 912-915.

40. Ashma R, Kashyap VK: **Genetic study of 15 important STR loci Among Four Major Ethnic Groups of Bihar, India**. *J. Forensic Sci*. 2002, **47:**1139-1142.

41. Rajkumar R, Kashyap VK: **Distribution of allele of 15 STR loci of the Powerplex™ 16 Multiplex in four predominant population groups of South India**. *Forensic Sci. Int*. 2002, **126**: 175-179.

42. Budowle B, Smith J, Moretti T, DiZinno J: ***DNA TYPING PROTOCOLS: Molecular Biology and Forensic Analysis.*** Edited by: McAndrews. C, Eaton Publishing. 2000.

43. Li CC: *First Course in Population Genetics*. Pacific Grove, CA,Boxwood, 1976.

44. Nei M, Tajima F, Tateno Y: **Accuracy of estimated phylogenetic trees from molecular data**. *J. Mol. Evol*. 1983, **19**: 153-170.

45. Cavalli-Sforza LL, Edwards AWF: **Phylogenetic Analysis: models and estimation procedures**. *Am J Hum Genet* 1967, **19**: 233-257

46. Shriver MD, Jin L, Boerwinkle E, Deka R, Ferrell R, Chakraborty R: **A novel measure of genetic distance for highly polymorphic tandem repeat loci**. *Mol. Biol. Evol*. 1995, **12**: 914-920.

47. Reynolds J, Weir BS, Cockerham CC: **Estimation of the coancestry coefficient: basis for a short-term genetic distance**. *Genetics* 1983, **105**: 767-779.

48. MICROSAT (http://lotka.stanford.edu/microsat.html)

49. Saitou N, Nei M: **The neighbor joining method: A new method for reconstructing phylogenetic trees**. *Mol. Biol. Evol*. 1987, **4**: 406-425.

50. Sneath PHA, Sokal RR: *Numerical Taxonomy*. W.H.Freeman, San Fransisco, 1973.

51. DISPAN (http://www.bio.psu.edu/IMEG)

52. Kumar S, Tamura K, Jakobsen IB, Nei M: *MEGA2: Molecular Evolutionary Genetics Analysis software*, Arizona State University, Tempe, Arizona, USA, 2001.

53. Bottleneck (http://www.ensam.infra.fr/URLB/bottleneck/bottleneck.html)

Table 1:- Heterozygosity and Gst (coefficient of gene differentiation) values at 15 microsatellite loci of the three populations of Sikkim.

| Locus \ Population | Heterozygosity | | | Gst |
|---|---|---|---|---|
| | Nepali | Bhutia | Lepcha | |
| D3S1358 | 0.666 | 0.750 | 0.728 | 0.00485 |
| THO1 | 0.715 | 0.657 | 0.432 | 0.01511 |
| D21S11 | 0.921 | 0.651 | 0.818 | 0.00994 |
| D18S51 | 0.842 | 0.843 | 0.811 | 0.01240 |
| D5S818 | 0.731 | 0.717 | 0.773 | 0.02489 |
| D13S317 | 0.777 | 0.688 | 0.750 | 0.01763 |
| D7S820 | 0.855 | 0.718 | 0.637 | 0.02232 |
| D16S539 | 0.742 | 0.875 | 0.613 | 0.00433 |
| CSF1PO | 0.742 | 0.750 | 0.660 | 0.02325 |
| VWA | 0.810 | 0.750 | 0.796 | 0.02642 |
| D8S1179 | 0.857 | 0.812 | 0.682 | 0.03654 |
| TPOX | 0.634 | 0.656 | 0.546 | 0.02418 |
| FGA | 0.800 | 0.563 | 0.523 | 0.08872 |
| PENTA E | 0.902 | 0.902 | 0.841 | 0.01943 |
| PENTA D | 0.804 | 0.875 | 0.660 | 0.02069 |
| Average | 0.786 | 0.747 | 0.684 | 0.0237 |

Table 2:- Shared allele distribution at 12 microsatellite loci of three studied populations of Sikkim along with comparative data from other populations.

| Locus | Nepali | Bhutia | Lepcha | Naga | Kuki | Hmar | Garo | Chinese | Caucasian | Brahmin (Karnataka) | Brahmin (Orissa) | Brahmin (Bihar) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| D3S1358 | 6 | 6 | 6 | 5 | 7 | 7 | 6 | 6 | 6 | 5 | 6 | 6 |
| THO1 | 4 | 6 | 4 | 5 | 5 | 5 | 5 | 5 | 4 | 4 | 5 | 5 |
| D21S11 | 15 | 10 | 16 | 9 | 13 | 13 | 6 | 10 | 10 | 12 | 11 | 10 |
| D18S51 | 15 | 11 | 11 | 10 | 11 | 14 | 13 | 13 | 10 | 13 | 11 | 12 |
| D5S818 | 6 | 6 | 6 | 7 | 8 | 10 | 10 | 8 | 6 | 6 | 5 | 5 |
| D13S317 | 6 | 6 | 7 | 6 | 7 | 7 | 6 | 8 | 7 | 7 | 8 | 8 |
| D7S820 | 8 | 6 | 7 | 7 | 6 | 7 | 9 | 7 | 6 | 7 | 7 | 7 |
| CSF1PO | 6 | 5 | 7 | 4 | 6 | 6 | 6 | 7 | 4 | 5 | 7 | 7 |
| VWA | 10 | 5 | 10 | 8 | 8 | 9 | 8 | 7 | 7 | 7 | 7 | 8 |
| D8S1179 | 10 | 9 | 9 | 9 | 8 | 9 | 9 | 10 | 8 | 9 | 11 | 9 |
| TPOX | 6 | 6 | 3 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 6 | 5 |
| FGA | 18 | 9 | 7 | 10 | 9 | 11 | 9 | 13 | 9 | 12 | 13 | 15 |

Table 3:- SIGN test for bottleneck effect in the three population of Sikkim**.**

| | Infinite Allele Model (IAM) | | | Stepwise Mutation Model (SMM) | | |
|---|---|---|---|---|---|---|
| Population | He<Heq | He>Heq | Probability | He<Heq | He>Heq | Probability |
| Nepali | 1 loci | 14 loci | 0.00484 | 7 loci | 8 loci | 0.41446 |
| Bhutia | 2 loci | 13 loci | 0.02380 | 10 loci | 5 loci | 0.04238 |
| Lepcha | 6 loci | 9 loci | 0.57428 | 13 loci | 2 loci | 0.00037 |

He<Heq : Heterozygosity deficiency, He>Heq : Heterozygosity excess

**Figure**

Figure 1:- Neighbor joining (NJ) phylogenies of the studied populations of Sikkim and other Indian and global populations.

Figure 2:- Probable route of migration of Mongoloids to Sikkim via Zang (Tibet) main corridor

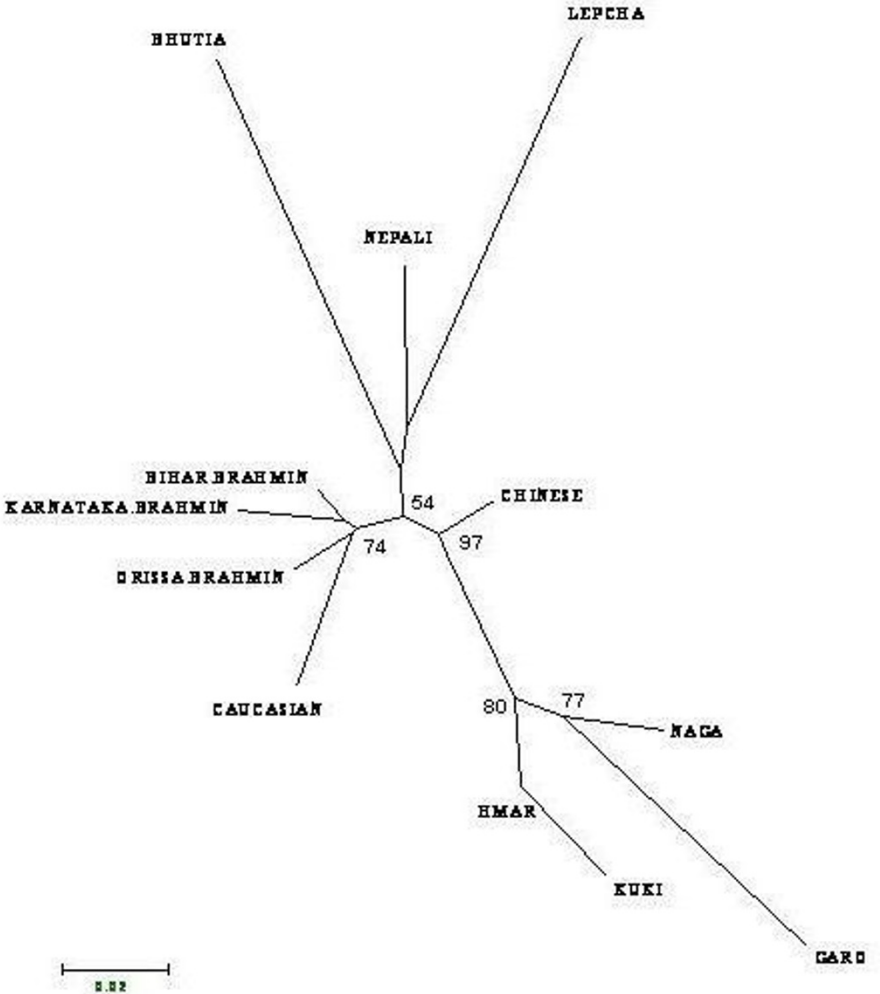
BHUTIA
LEPCHA
NEPALI
BIHAR BRAHMIN
KARNATAKA BRAHMIN
ORISSA BRAHMIN
CAUCASIAN
CHINESE
NAGA
HMAR
KUKI
GARO
54
74
97
80
77
0.02

Figure 1


JAMMU & KASHMIR
Main route of migration
N
TIBBET
CHINA
PAKISTAN
UTTARANCHAL
Migration to Sikkim
DELHI
NEPAL
SIKKIM
BHUTAN
Gangtok
Patna
JHARKHAND
CHHATTISGARH
Calcutta
Mumbai (Bombay)
ARABIAN SEA
BAY OF BENGAL

Figure 2