

Swarkar Sharma · Anjana Saha · Ekta Rai
Audesh Bhat · Ramesh Bamezai

Human mtDNA hypervariable regions, HVR I and II, hint at deep common maternal founder and subsequent maternal gene flow in Indian population groups

Received: 6 May 2005 / Accepted: 21 July 2005 / Published online: 5 October 2005
© The Japan Society of Human Genetics and Springer-Verlag 2005

Abstract We have analysed the hypervariable regions (HVR I and II) of human mitochondrial DNA (mtDNA) in individuals from Uttar Pradesh (UP), Bihar (BI) and Punjab (PUNJ), belonging to the Indo-European linguistic group, and from South India (SI), that have their linguistic roots in Dravidian language. Our analysis revealed the presence of known and novel mutations in both hypervariable regions in the studied population groups. Median joining network analyses based on mtDNA showed extensive overlap in mtDNA lineages despite the extensive cultural and linguistic diversity. MDS plot analysis based on F_{st} distances suggested increased maternal genetic proximity for the studied population groups compared with other world populations. Mismatch distribution curves, respective neighbour joining trees and other statistical analyses showed that there were significant expansions. The study revealed an ancient common ancestry for the studied population groups, most probably through common founder female lineage(s), and also indicated that human migrations occurred (maybe across and within the Indian subcontinent) even after the initial phase of female migration to India.

Keywords Indian populations · Mitochondrial DNA · Matrilineal · HVR I and II

Introduction

The distribution of human genetic diversity has long been a subject of interest, and it has important implications for human evolution, forensics, and the distribution of genetic diseases in populations. Genetic

diversity in human populations is low relative to that in many other species, attesting to the recent origin and small size of the ancestral human population (Li and Sadler 1991; Crouau-Roy et al 1996; Kaessmann et al 1999). Since the seminal study of Cann et al (1987), mitochondrial DNA (mtDNA) data have proven to be extremely useful for studying human evolution, including prehistoric migrations and demographic events such as sudden population expansions or extreme bottlenecks (Sherry et al 1994).

The human mtDNA is a closed circular genome of ~16.5 kb in length (Anderson et al 1981), which includes a 1.1 kb-long noncoding (control) region that represents a highly variable sequence (Greenberg et al 1983; Melton et al 1997). The variable sites in the control regions, hypervariable region I (HVR I) and hypervariable region II (HVR II), each ~400 bp, correspond to the origin of replication and the D-loop. Since mtDNA is maternally inherited, exists in a high copy number (1,000–10,000 copies) in each cell, and is rapidly evolving (5–10 times faster than nuclear DNA), sequence polymorphism of the mtDNA has proved useful in the fields of population and evolution study (Gresham et al 2001; Ingman and Gyllenstein 2001; Roychoudhury et al 2001; Yao et al 2002; Kivisild et al 1999; Metspalu et al 2004; Forster and Matsumura 2005; Thangaraj et al 2005; Macaulay et al 2005), anthropology (Derbeneva et al 2002; Houck and Budowle 2002; Koyama et al 2002; Yao and Zhang 2002) and forensic science (Budowle et al 2002; Seo et al 2002).

India comprises one of the largest ethnic populations, with more than 1 billion people drawn from diverse cultures, languages and geographical backgrounds. A number of studies have provided some insights into the maternal genetic structure of Indian populations (Bamshad et al 2001; Kaur et al 2002; Basu et al 2003; Kivisild et al 2003; Rajkumar and Kashyap 2003; Metspalu et al 2004; Palanichamy et al 2004; Rajkumar et al 2005; Thangaraj et al 2005). However, more studies are required to add to the pool of information and to help us to better understand the

S. Sharma · A. Saha · E. Rai · A. Bhat · R. Bamezai (✉)
National Centre of Applied Human Genetics,
School of Life Sciences, Jawaharlal Nehru University,
110067 New Delhi, India
E-mail: bamezai@hotmail.com
Tel.: +91-11-2613211

genetic structures of diverse Indian population groups, where many questions remain unanswered. The present preliminary work proposed to analyse the nature of the variations in hypervariable regions (HVR I and II) and perform phylogenetic analyses on individuals from Uttar Pradesh (UP), Bihar (BI) and Punjab (PUNJ), belonging to the Indo-European linguistic group, and individuals from South India (SI), that have their linguistic roots in the Dravidian language, in order to derive the structures of founder female lineages for these regions in India. We have also focused on providing an overview of the maternal gene pool, regional maternal population expansion time/patterns and their phylogenetic relationships with each other and other world populations.

Materials and methods

Subjects

Blood samples were collected (after seeking the required consent) in RBC lysis buffer from unrelated healthy donors from different population groups of India. One hundred eleven samples were collected, which belonged to: UP ($n=33$), BI ($n=26$), PUNJ ($n=35$) and SI ($n=17$). Genomic DNA was isolated following the routine protocol of Kunkel et al (1977).

Amplification of mtDNA

PCR amplifications of mtDNA regions—HVR I, for a total of 111 individuals, by a designed primer set AB-6F (5'-ACC CAA TCC ACA TCA AAA CC-3') and AB-6R (5'-TCA AGG GAC CCC TAT CTG AG-3') and HVR-II, for 87 individuals, by designed primer set F (5'-GGT CTA TCA CCC TAT TAA CCA C-3') and R (5'-CTG TTA AAA GTG CAT ACC GCC A-3')—were performed in 12.5 μ l reaction volume mix containing 50 ng of template DNA, 6.25 pmol of each primer, 200 μ M of dNTPs, 1.5 mM $MgCl_2$, 1 \times reaction buffer and 0.3 U of *Taq pol* enzyme (Bangalore Genei, India). The cycle used was denaturation at 94 °C for 1 min, followed by annealing at 62 °C for 1 min, and then extension at 72 °C for 1 min, repeated for 30 cycles followed by a final extension step at 72 °C for 5 min. PCR products were initially checked in 2% agarose gel and then sequenced (using an ABI 3100 sequencer, USA). The sequences obtained were compared with the revised Cambridge Reference Sequence (CRS) (Anderson et al 1981) to find mutations.

Phylogenetic and statistical analysis

The regional/linguistic descriptions and HVR I haplotypic motifs of the samples are given in Appendix 1. The haplotypic motifs were defined as putative haplogroups by comparing them with HVR I data obtained from

Metspalu et al (2004), and the coalescence ages for these haplogroups and their standard errors (SE) for mutation rates were calculated as described by Forster et al (1996). Mitochondrial hypervariable regions (HVR I and II) were analysed for statistical and phylogenetic patterns. The software DNASP 4.0.0.4 (Rozas and Rozas 1999) was used to identify the number of polymorphic sites and number of mutations, to calculate nucleotide diversity, the mean number of mismatches, Fu's F_s statistics (Fu 1997), Tajima's D values (Tajima 1989), raggedness statistics and to draw graphical patterns of mismatch distributions (Rogers and Harpending 1992). The other statistical inferences, like initial theta (θ_a) and values of tau (τ), obtained from DNASP 4.0.0.4, were used to calculate N_e = effective population size ($\theta_a/2 \mu$) and population expansion age $^AY_a = (A \times \tau/2 \mu)$ years ago (Rogers and Harpending 1992). An average mutation rate $\mu = 0.00124$ per site per generation (Forster et al 1996) with an average generation time $A = 20$ years, was used for the calculations.

The neighbour-joining (NJ) trees for the studied population groups were generated using MEGA 2.1 software. Pairwise genetic distances between studied populations were computed as a linearisation of FST/(1-FST) (Slatkin 1995) using ARLEQUIN 2.000 software (Schneider et al 2000). These linearised distance values were used to create a MDS plot based on HVR I data using SPSS 10.0.5 (Chicago, IL, USA) and a NJ Tree, based on HVR II data, using PHYLIP (Version 3.5c) (Felsenstein 1989; PHYLIP home page) and a TreeView (version 1.6.1) (see <http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>) for the studied population groups. Median-joining (MJ) networks (Bandelt et al 1999) for HVR I and II were constructed using the software NETWORK 4.1.0.8.

Results

Analysis of mitochondrial HVR I (nucleotide positions: 16023–16414) and HVR II (nucleotide positions: 53–424) showed the presence of a large number of variations, both known and novel mutations, in the samples analysed. Some of the sequences with mutations were deposited in the GenBank database and accession numbers were obtained: HVR I (AF467445–450, AF542192–198 and AY899182–94) and HVR II (AY642000–023).

The analysis of the putative haplogroups classified on the basis of mtDNA HVR I haplotypic motifs (Appendix 1) showed that 53.15% of the individuals from the studied population groups belonged to haplogroup M lineages with a coalescence age of $66,043.6 \pm 9,995.7$ years. Both U2 and U7 lineages of haplogroup U, each 10.8% of the total, showed a deep but different coalescence age of $58,297.7 \pm 23,729$ and $35,875 \pm 10,984.6$ years, respectively. Individuals represented in haplogroup R lineages (9%), most of them (5.4%) belonging to R5, showed a coalescence age of $52,468 \pm$

17592.5 years. Other observed lineages, known to be of European origin (F, H, HV, J, U1 and U5), represented 9.9% of the total population, whereas 6.3% remained undefined.

Using the HVR I sequence data, the number of polymorphic sites, the number of mutations, the mean number of mismatches, the nucleotide diversities, the initial theta values, the raggedness statistic values, Fu's F_s statistic values, Tajima's D values, expansion ages and initial effective population sizes of all the studied population groups were calculated and are given in Table 1. Figure 1 depicts the mismatch distribution patterns and respective NJ trees for the studied population groups based on mitochondrial HVR I data. The unimodal nature, the smoothness (as revealed by the very small values for the raggedness statistics) of the mismatch distribution curves, the reasonably good fits with the expected distributions of the observed mismatch distributions, the branching patterns in the NJ trees, the significantly large negative values of Fu's F_s statistics, and the highly significant values of Tajima's D (Table 1) clearly indicate that there were significant expansions of the different population groups studied, which is also supported by the star-like configurations in the MJ networks based on HVR I (Fig. 2a) and HVR II (Fig. 2b) data. An extreme sharing of haplotypes with no population-specific differentiation was also observed in both of the MJ networks.

An MDS plot based on the genetic distances of Slatkin linearised FSTs from mtDNA HVR I data (Fig. 3a) showed that the studied population groups form a compact group and this cluster includes Mongol, Egyptian and sub-Saharan as the nearest population groups. An NJ Tree (Fig. 3b) based on the genetic distances of Slatkin linearised FSTs (given as UP and BI = 0.00, UP and SI = 0.00, BI and SI = 0.00, PUNJ and BI = 0.05, PUNJ and SI = 0.04, PUNJ and UP = 0.09) from HVR II data revealed that the population groups UP and BI formed a single cluster whereas PUNJ branched out, probably depicting genetic affinity among the UP, BI and SI population groups as compared to PUNJ.

Discussion

The geography of India has played a decisive role in the peopling of India. Populations within India have been

subjected to foreign invasions and migrations from time to time, resulting in no single apparent origin for any present day population groups and a conglomeration of different Y-chromosomal lineages (Quintana-Murci et al 2001; Saha et al 2005). The maternal gene flow in and out of India has been limited since the initial settling of Indian maternal lineages (Metspalu et al 2004). Indian mtDNA lineages belong to either Asian-specific haplogroup M or western Eurasian-specific haplogroups H, I, J, K, U, W and others that were not established anywhere (Kivisild et al 1999). The high frequency and diversity of mtDNA haplogroup M, the major contributor to the Indian maternal gene pool, has been associated with its southwest-Asian origin (Roychoudhury et al 2000, 2001; Richards et al. 2003; Rajkumar et al. 2005), whereas the presence of lineage M1 in Africa (Quintana-Murci et al 1999) and lack of L3 lineages other than M and N in India has become the most parsimonious view of the origin of haplogroup M in east Africa, which has been supported by the most recent view of single rapid coastal settlement of Asia by three major mtDNA haplogroups, M, N and R (Palanichamy et al 2004; Macaulay et al 2005; Thangaraj et al 2005; Forster and Matsumura 2005) as the founding female lineages to Indian population groups. However, the restricted presence of M as M1 and the phylogeography of M1 in Africa, predominantly in the Afro-Asiatic linguistic phylum (Metspalu et al 2004), leaves the question of the origin of haplogroup M unanswered.

This study of human mtDNA attempts to answer a variety of questions regarding the structures and compositions, the genetic relationships, the ancestries and the effects of migrations on present day females within the population groups studied in India, and it also throws light on their genetic relationships with each other and with other world populations. The observed deep coalescence ages, distribution patterns and high diversities of haplogroup M in all of the studied population groups supported the concept of a common founder of the Indian maternal gene pool acting as a major contributor, irrespective of its place of origin. Further, the overlapping coalescence ages within the standard deviation intervals for the U2 and R5 lineages suggested their coexistence with haplogroup M lineages (Metspalu et al 2004), but comparatively less so in frequency and diversity, which could also be due to independent migration events in India. Whereas, the coalescence age of U7 was observed to

Table 1 Descriptive statistics based on HVR I in Indian population groups belonging to various linguistic backgrounds

Pop. Group	N_a	N_b	N_c	π (SD)	k	τ	R	θ_a	4Y_a	N_e	F_s	D
UP	33	57	57	0.017 (0.001)	6.542	6.297	0.010	0.245	50766.13	98.790	-29.115	-1.972*
BI	26	52	55	0.018 (0.001)	7.132	7.132	0.015	0	57516.13	0	-20.269	-1.939*
PUNJ	35	51	52	0.016 (0.002)	6.185	4.668	0.010	1.516	37645.16	611.29	-30.295	-1.858*
SI	17	46	46	0.019 (0.002)	7.588	7.588	0.027	0	61193.55	0	-8.372	-1.840*

N_a = number of sequences; N_b = number of polymorphic sites; N_c = number of mutations; π (SD) = nucleotide diversity (SD); k = mean number of mismatches; τ = tau; θ_a = initial theta; r = raggedness; F_s = Fu's F_s statistic; $D(p)$ = Tajima's D ; * = $P < 0.05$. 4Y_a = years ago ($A \times \tau/2 \mu$); N_e = effective population size ($\theta_a/2 \mu$); μ = mutation rate (0.00124 per site per generation); A = generation time (20 years)

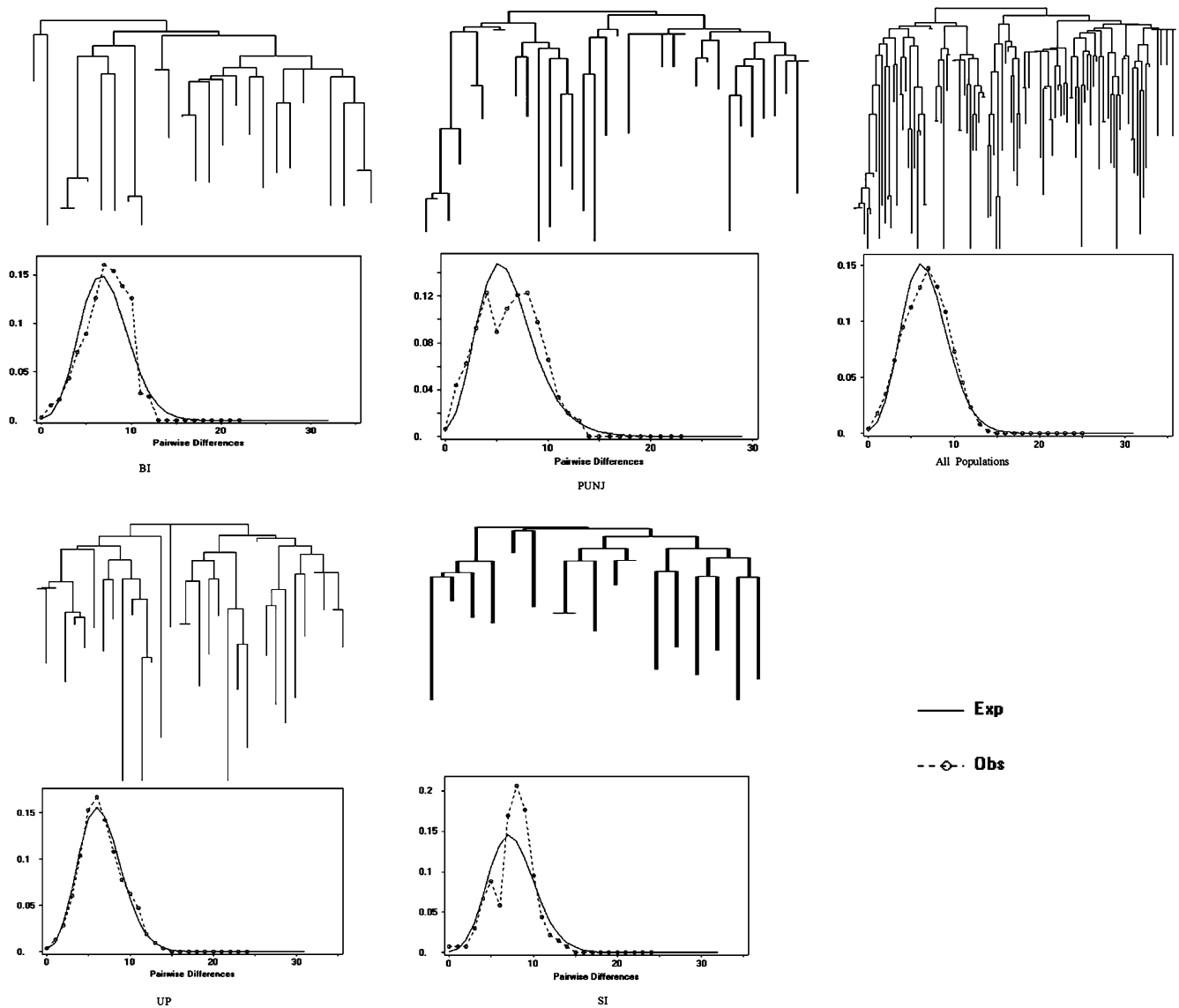


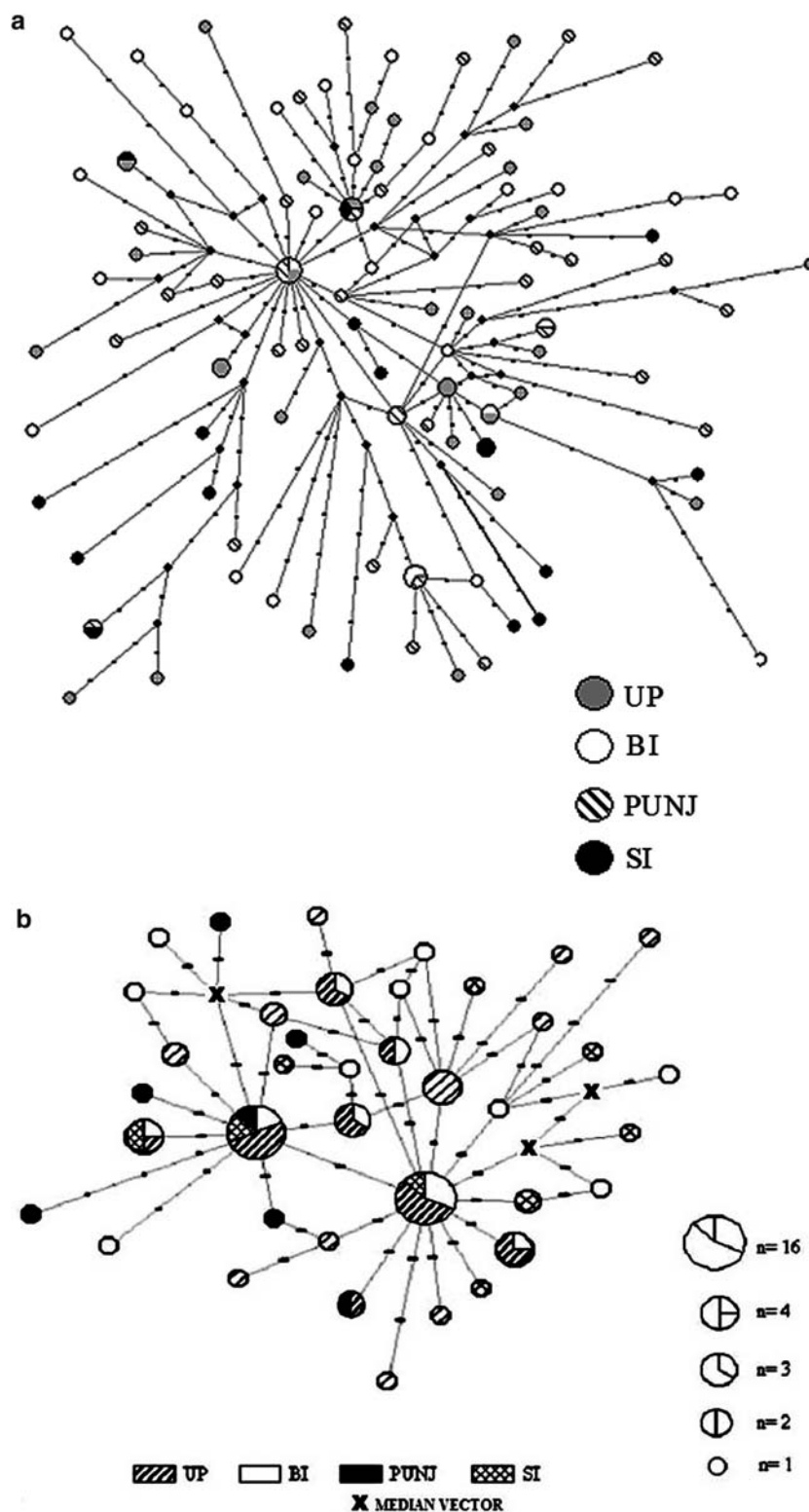
Fig. 1 Observed (*dashed line*) and expected (*solid line*) mismatch distribution curves and respective neighbour-joining trees showing population expansion patterns based on mtDNA HVR I data

be far younger, indicating that it was differentiated later and that the trans-Indian subcontinent spread occurred later too. The mismatch distribution patterns, respective NJ trees, various statistical analyses (Table 1), star-like configurations of the clusters with extreme sharing and the lack of population-specific differentiation in the MJ network (Fig. 2a, b) suggested that all of the population groups underwent expansions in ancient times and had a fundamental maternal similarity. There were also indications of some recent mtDNA migrations, apparent in the form of some small clusters without star-shaped phylogeny in the MJ network (Fig. 2a), that could have been associated with various recent invasions of and migration events to India. A greater maternal genetic proximity was revealed in MDS plot analysis (Fig. 3a) for the studied population groups when compared with other world populations, which also indicated the conservation

of the east-Asian mtDNA components and the presence of West Eurasian and African/sub-Saharan lineages. The clustering in the NJ tree (Fig. 3b) and the F_{st} values indicated a close maternal relationship between all of the Indian populations, whereas the branching out of the Punjab supports the interpretation that there was probably a considerable inflow of genes from Indo-European-speaking populations from central and possibly from West Asia into the Punjab (Passarino et al 1996; Kaur et al 2002). The results obtained also suggest that linguistic/ethnic differences evolved later on, by the process of acculturation, and the recent demic diffusion (Quintana-Murci et al 2001) also brought in some western-Eurasian mtDNA components.

To conclude, the present study supports an ancient common ancestry for the studied population groups through common founder female lineages, but it also

Fig. 2a,b **a** Median-joining network of the studied populations based on mtDNA HVR I haplotypes. **b** Median-joining network of the studied populations based on mtDNA HVR II haplotypes



indicates a maternal gene flow in ancient times with further ethnic differentiation occurring subsequently through a series of demographic expansions, geographical dispersals, social groupings and later Eurasian admixture.

In future studies, a larger sample size and mitochondrial coding region SNP markers could provide more information and lead to a better comprehension of the evolutionary history of the studied and other Indian population groups.

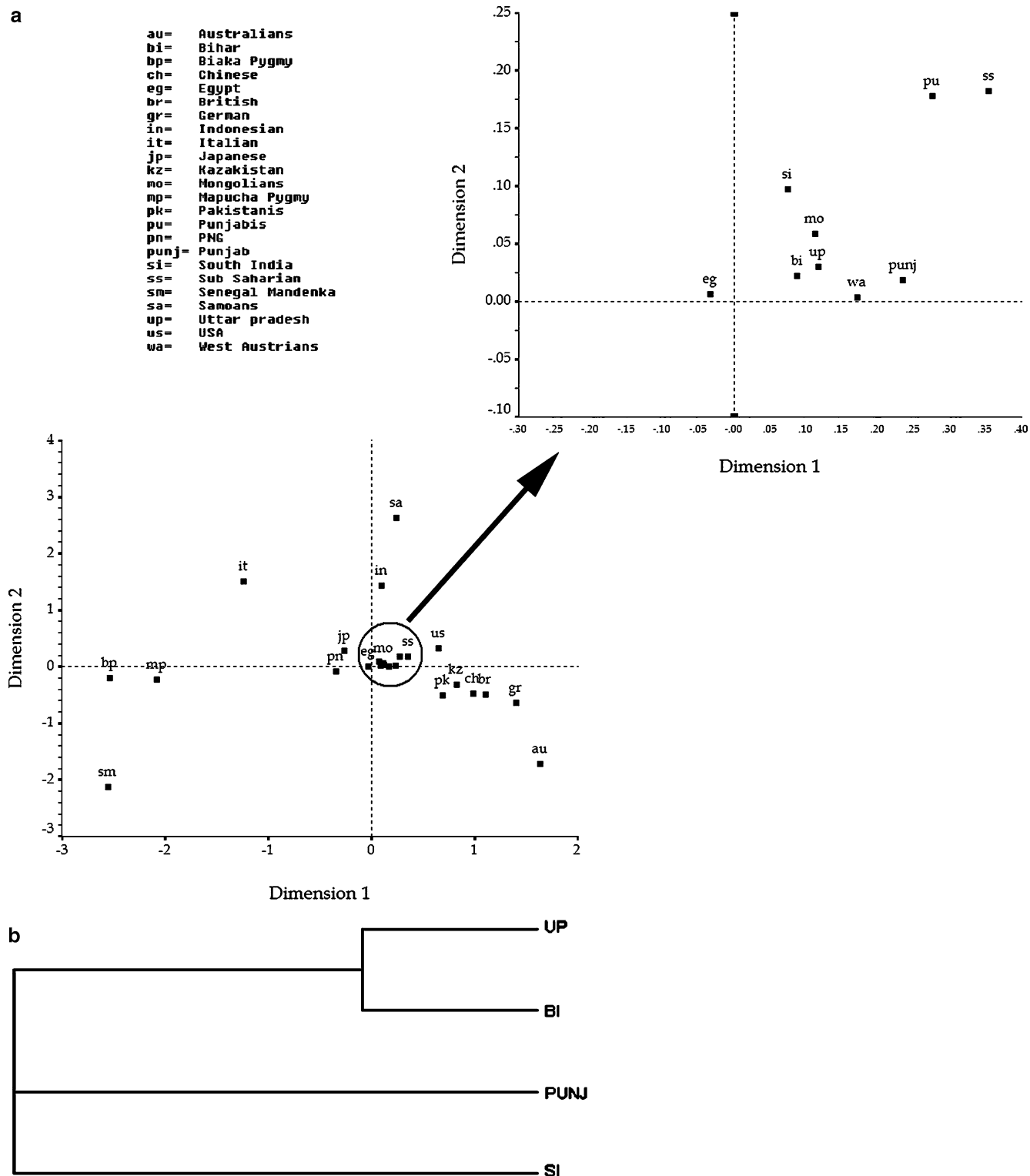


Fig. 3a,b a MDS plot based on Slatkin's linearised F_{st} values obtained from HVR I data showing the genetic relationship between the studied and other world population groups.

b Neighbour-joining tree of the studied population groups based on the genetic distances of Slatkin linearised F_{ST} s obtained from HVR II data

Appendix

Appendix 1 Regional/linguistic descriptions and HVR I haplotypic motifs of the studied samples

S. no	Sample	Region	Linguistic group	mtDNA region (16,023–16,414); HVS I motif (-16000)	Putative haplogroup ^a
1	B16	Bihar	Indo-European	129, 144A, 148G, ^b 223, 367T ^b	M/M5
2	B37	Bihar	Indo-European	223, 302	M
3	BD10	Bihar	Indo-European	93, 129, 145, 182C, 183C, 189, 318T	U/U1/U7
4	BD13	Bihar	Indo-European	51, 93A, 154, 206C, 230, 311	U2a
5	BD113	Bihar	Indo-European	223, 234	M
6	BD14	Bihar	Indo-European	126, 223, 368, 391	M3
7	BD27	Bihar	Indo-European	129, 223, 362	M5
8	BD28	Bihar	Indo-European	145, 176G, 223, 240, 390	N.D.
9	BD31	Bihar	Indo-European	223, 234, 305, 311, 362	M9a
10	BD33	Bihar	Indo-European	126, 223, 301, 344	M3
11	BD37	Bihar	Indo-European	51, 192, 287, 325, 381	U2b
12	BD39	Bihar	Indo-European	309, 318T	U7
13	BD52	Bihar	Indo-European	86, 223, 284, 327, 398	M
14	BD55	Bihar	Indo-European	209, 239, 352, 353	N.D.
15	BHC113	Bihar	Indo-European	172, 304, 362	F/R
16	BHC115	Bihar	Indo-European	362	MD4
17	BHC130	Bihar	Indo-European	223	L3/M/N
18	BHC134	Bihar	Indo-European	51, 209, 239, 352, 353	U2b
19	BHC14	Bihar	Indo-European	129, 223, 265C, 311	M5
20	BHC156	Bihar	Indo-European	223, 275, 304G, 327A, ^b 390	M
21	BHC27	Bihar	Indo-European	51, 104, 207, 227, 408 ^b	U2
22	BHC56	Bihar	Indo-European	223, 275, 327A, ^b 390	M
23	BHC66	Bihar	Indo-European	75, 93, 129, 223, 291	M
24	BHC91	Bihar	Indo-European	129, 169, 223, 391	M
25	BSUY	Bihar	Indo-European	51, 209, 239, 352, 353	U2b
26	BVIJ	Bihar	Indo-European	129, 223, 291	M5
27	P14	Punjab	Indo-European	51, 209, 239, 352, 353	U2b
28	PBAN	Punjab	Indo-European	166delA, 223	M
29	PHC50	Punjab	Indo-European	93, 266, 325, 356	R5
30	PHC68	Punjab	Indo-European	126, 223, 266, 311	M3a
31	PNIR	Punjab	Indo-European	223	L3/M/N
32	PP55	Punjab	Indo-European	145, 176, 223, 261, 266, 291, 311	M4a
33	PP56	Punjab	Indo-European	172, 304, 362	F/R
34	PP57	Punjab	Indo-European	172, 318T	U7
35	PP59	Punjab	Indo-European	126, 223, 260	M3
36	PP62	Punjab	Indo-European	129, 223, 265C	M5
37	PP63	Punjab	Indo-European	223, 362	M/MD/MD4
38	PP64	Punjab	Indo-European	179delC, ^b 223	M
39	PP65	Punjab	Indo-European	69, 126, 145, 222, 261, 288, 362	J
40	PP67	Punjab	Indo-European	223, 294, 311, 318T	U7/M4
41	PP68	Punjab	Indo-European	126, 129, 169, 223	M3a
42	PP69	Punjab	Indo-European	182, 311, 408 ^b	M (N.D.)
43	PP70	Punjab	Indo-European	129, 148, 266, 290, 318, ^b 320, 362	R6
44	PP76	Punjab	Indo-European	129, 214A, ^b 223, 252, 291, 298	M/M5
45	PP77	Punjab	Indo-European	182, 311, 223, 304	M25 (M- N.D.)
46	PP83	Punjab	Indo-European	223, 304	M/M25
47	PP84	Punjab	Indo-European	223, 293	M
48	PP86	Punjab	Indo-European	129, 223, 296, 311	M5
49	PP87	Punjab	Indo-European	126, 173, 181, 209, 260, 362	R21
50	PP88	Punjab	Indo-European	51, 209, 239, 311, 352, 353	U2b
51	PP89	Punjab	Indo-European	193, 218, 223, 335	M (N.D.)
52	PP91	Punjab	Indo-European	126, 223, 304	M3/M25
53	PP94	Punjab	Indo-European	182, 311, 408 ^b	M (N.D.)
54	PP97	Punjab	Indo-European	69, 126, 261	J
55	PPUR	Punjab	Indo-European	129, 223	M/M5
56	PPUS	Punjab	Indo-European	93, 192, 256, 270, 362, 399	U5
57	PRAJ	Punjab	Indo-European	172, 256, 266, 304, 356	R5
58	PRAM	Punjab	Indo-European	51, 239, 288, 353	U2b
59	PSUR	Punjab	Indo-European	51, 209, 239, 246, 255delG, ^b 352, 353	U2b
60	PSUS	Punjab	Indo-European	223, 257, 261, 324, 362	M
61	PUSH	Punjab	Indo-European	223, 234	M
62	S12	South India	Dravidian	126, 223, 278, 327	M3a
63	S18	South India	Dravidian	51, 209, 239, 244, ^b 352, 353	U2b
64	S22	South India	Dravidian	129, 218, 309, 318T	U7

Appendix 1 (Continued)

65	S28	South India	Dravidian	37, 148G, 215, 390	N.D.
66	SHC13	South India	Dravidian	192, 291, 295, 298, 390	N.D.
67	SHC15	South India	Dravidian	183C, 189, 210, 223, 266	M
68	SHC198	South India	Dravidian	69, 274, 318T	U7
69	SHC51	South India	Dravidian	129, 223	M/M5
70	SHC53	South India	Dravidian	51, 86, 259, 261, 291, 353	U2b
71	SHC98	South India	Dravidian	69, 274, 318T	U7
72	SHC99	South India	Dravidian	256, 270, 311, 399	U5
73	SR02	South India	Dravidian	172, 256, 266, 304, 356	R5
74	SST14	South India	Dravidian	93, 148G, ^b 210, 223, 266	M/N.D.
75	SST17	South India	Dravidian	223, 318T	M18
76	SST3	South India	Dravidian	93, 124, 223, 266, 349 ^b	M
77	SST4	South India	Dravidian	93, 100, ^b 140A, ^b 194, 223, 233, 241T, ^b 266	R5/N.D.
78	SST7	South India	Dravidian	223, 318T, 322	M18
79	U15	Uttar Pradesh	Indo-European	92, 287	N.D.
80	U68	Uttar Pradesh	Indo-European	66C, ^b 126, 145, 176, 223, 261, 311	M4a
81	UANI	Uttar Pradesh	Indo-European	129, 223, 274, 297, 311, 362	MD4/MG3/M5
82	UBHA	Uttar Pradesh	Indo-European	126, 223, 408 ^b	M3
83	UBOB	Uttar Pradesh	Indo-European	51, 223, 249, 250	N.D.
84	UCHA	Uttar Pradesh	Indo-European	223, 234, 292, 311, 316, 362	M/M9
85	UD113	Uttar Pradesh	Indo-European	37, 189, 318T	U7
86	UD18	Uttar Pradesh	Indo-European	309, 318T, 362	U7
87	UD24	Uttar Pradesh	Indo-European	318T	U7
88	UP124	Uttar Pradesh	Indo-European	318T	U7
89	UD43	Uttar Pradesh	Indo-European	69, 126, 145, 172, 222, 261, 355	J
90	UD47	Uttar Pradesh	Indo-European	54, ^b 266, 304	R5
91	UD50	Uttar Pradesh	Indo-European	51, 65T, ^b 66C, ^b 129, 172, 206C	N.D./U2
92	UD56	Uttar Pradesh	Indo-European	309, 318T	U7
93	UD64	Uttar Pradesh	Indo-European	93, 129, 223, 291	M/M5
94	UHC11	Uttar Pradesh	Indo-European	126, 223, 278, 327	M3a
95	UHC114	Uttar Pradesh	Indo-European	176, 223, 344	M
96	UHC116	Uttar Pradesh	Indo-European	304, 311	HV3/R
97	UHC126	Uttar Pradesh	Indo-European	129, 223	M/M5
98	UHC128	Uttar Pradesh	Indo-European	54, ^b 266, 304, 311, 355, 356, 399	R5
99	UHC139	Uttar Pradesh	Indo-European	93, 223, 225, ^b 234, 290	R
100	UHC2	Uttar Pradesh	Indo-European	42, 66T, ^b 129, 223, 243	M/M5
101	UHC24	Uttar Pradesh	Indo-European	129, 223, 243	M/M5
102	UHC214	Uttar Pradesh	Indo-European	176, 223, 344	M
103	UHC31	Uttar Pradesh	Indo-European	86, 145, 223, 261, ^b 311	M4a
104	UHC32	Uttar Pradesh	Indo-European	48, 129, 223	M/M5
105	UHC64	Uttar Pradesh	Indo-European	126, 181, 223, 233, 344	R21/N.D.
106	UHC70	Uttar Pradesh	Indo-European	223, 213, 362	M6a
107	UHC75	Uttar Pradesh	Indo-European	51, 209, 239, 244, ^b 304, 352, 353	U2b
108	UJEA	Uttar Pradesh	Indo-European	124, 223, 274, 362	MD4/MG3
109	UKIR	Uttar Pradesh	Indo-European	309, 318T, 368	U7
110	USIN	Uttar Pradesh	Indo-European	304, 320, 362	N.D./H
111	63	Uttar Pradesh	Indo-European	222delC, 225	N.D./H

^aPutative haplogroups defined by HVS I motifs; N.D. = not defined; / = to be confirmed by mtDNA coding region polymorphism

^bNovel mutations observed when compared with control region polymorphisms database (source: Brandon et al 2005)

Acknowledgement The financial assistance to SS and ER as NCAHG (UGC) fellows and AS in the form of a Research Associateship by CSIR is acknowledged. The authors also acknowledge the grant given to the National Centre of Applied Human Genetics by UGC, India, and the support as a Centre for Advanced Studies to the School of Life Sciences.

References

- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465
- Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BV, Reddy PG, Rasanayagam A, Papiha SS, Vilems R, Redd AJ, Hammer MF, Nguyen SV, Carroll ML, Batzer MA, Jorde LB (2001) Genetic evidence on the origins of Indian caste populations. *Genome Res* 11:994–1004
- Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48
- Basu A, Mukherjee N, Roy S, Sengupta S, Banerjee S, Chakraborty M, Dey B, Roy M, Roy B, Bhattacharyya NP, Roychoudhury S, Majumder PP (2003) Ethnic India: a genomic view, with special reference to peopling and structure. *Genome Res* 13:2277–2290
- Brandon MC, Lott MT, Nguyen KC, Spolim S, Navathe SB, Baldi P, Wallace DC (2005) MITOMAP: a human mitochondrial genome database: 2004 update. *Nucleic Acids Res* 33:D611–D613 (see also <http://www.mitomap.org>)

- Budowle B, Allard MW, Wilson MR (2002) Critique of interpretation of high levels of heteroplasmy in the human mitochondrial DNA hypervariable region I from hair. *Forensic Sci Int* 126:30–33
- Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325:31–36
- Crouau-Roy B, Service S, Slatkin M, Freimer N (1996) A fine-scale comparison of the human and chimpanzee genomes: linkage, linkage disequilibrium and sequence analysis. *Hum Mol Genet* 5:1131–1137
- Derbeneva OA, Sukernik RI, Volodko NV, Hosseini SH, Lott MT, Wallace DC (2002) Analysis of mitochondrial DNA diversity in the Aleuts of the Commander Islands and its implications for the genetic history of Beringia. *Am J Hum Genet* 71:415–421
- Felsenstein J (1989) PHYLIP—Phylogeny Inference Package (version 3.2). *Cladistics* 5:164–166
- Forster P, Harding R, Torroni A, Bandelt HJ (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 59:935–945
- Forster P, Matsumura S (2005) Evolution: Did early humans go north or south? *Science* 308:965–966
- Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915–925
- Greenberg BD, Newbold JE, Sugino A (1983) Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene* 21:33–49
- Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise C, Angelicheva D, Calafell F, Oefner PJ, Shen P, Tournev I, de Pablo R, Kucinkas V, Perez-Lezaun A, Marushiakova E, Popov V, Kalaydjieva L (2001) Origins and divergence of the Roma (gypsies). *Am J Hum Genet* 69:1314–1331
- Houck MM, Budowle B (2002) Correlation of microscopic and mitochondrial DNA hair comparisons. *J Forensic Sci* 47:964–967
- Ingman M, Gyllenstein U (2001) Analysis of the complete human mtDNA genome: methodology and inferences for human evolution. *J Hered* 92:454–461
- Kaessmann H, Wiebe V, Paabo S (1999) Extensive nuclear DNA sequence diversity among chimpanzees. *Science* 286:1159–1162
- Kaur I, Roy S, Chakrabarti S, Sarhadi VK, Majumder PP, Bhanwer AJ, Singh JR (2002) Genomic diversities and affinities among four endogamous groups of Punjab (India) based on autosomal and mitochondrial DNA polymorphisms. *Hum Biol* 74:819–836
- Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, Ferak V, Villems R (1999) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol* 9:1331–1334
- Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V, Golge M, Usanga E, Papiha SS, Cinnioglu C, King R, Cavalli-Sforza L, Underhill PA, Villems R (2003) The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet* 72:313–332
- Koyama H, Iwasa M, Ohtani S, Ohira H, Tsuchimochi T, Maeno Y, Isobe I, Matsumoto T, Yamada Y, Nagao M (2002) Personal identification from human remains by mitochondrial DNA sequencing. *Am J Forensic Med Pathol* 23:272–276
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Arizona State University, Tempe, AZ
- Kunkel LM, Smith KD, Boyer SH, Bargaonkar DS (1977) Analysis of human Y chromosome specific reiterated DNA in chromosome variants. *P Natl Acad Sci USA* 74:1245–1249
- Li WH, Sadler LA (1991) Low nucleotide diversity in man. *Genetics* 129:513–523
- Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, Taha A, Shaari NK, Raja JM, Ismail P, Zainuddin Z, Goodwin W, Bulbeck D, Bandelt HJ, Oppenheimer S, Torroni A, Richards M (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308:1034–1036
- Melton T, Wilson M, Batzer M, Stoneking M (1997) Extent of heterogeneity in mitochondrial DNA of European populations. *J Forensic Sci* 42:437–446
- Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MT, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, Villems R (2004) Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet* 5:26
- Palanichamy MG, Sun C, Agrawal S, Bandelt HJ, Kong QP, Khan F, Wang CY, Chaudhuri TK, Palla V, Zhang YP (2004) Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *Am J Hum Genet* 75:966–978
- Passarino G, Semino O, Bernini LF, Santachiara-Benerecetti AS (1996) Pre-caucasoid and caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms. *Am J Hum Genet* 59: 927–934
- Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat Genet* 23:437–441
- Quintana-Murci L, Krausz C, Zerjal T, Sayar SH, Hammer MF, Mehdi SQ, Ayub Q, Qamar R, Mohyuddin A, Radhakrishna U, Jobling MA, Tyler-Smith C, McElreavey K (2001) Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am J Hum Genet* 68:537–542
- Rajkumar R, Kashyap VK (2003) Haplotype diversity in mitochondrial DNA hypervariable regions I and II in three communities of Southern India. *Forensic Sci Int* 136:79–82
- Rajkumar R, Banerjee J, Gunturi HB, Trivedi R, Kashyap VK (2005) Phylogeny and antiquity of M macrohaplogroup inferred from complete mtDNA sequence of Indian specific lineages. *BMC Evol Biol* 5:26
- Richards M, Rengo C, Cruciani F, Gratrix F, Wilson JF, Scozzari R, Macaulay V, Torroni A (2003) Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. *Am J Hum Genet* 72:1058–1064
- Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 9:552–569
- Roychoudhury S, Roy S, Dey B, Chakraborty M, Roy M, Roy B, Ramesh A, Prabhakaran N, Rani MVU, Vishwanathan MM (2000) Fundamental genomic unity of ethnic India is revealed by analysis of mitochondrial DNA. *Curr Sci* 79:1182–1192
- Roychoudhury S, Roy S, Basu A, Banerjee R, Vishwanathan H, Usha Rani MV, Sil SK, Mitra M, Majumder PP (2001) Genomic structures and population histories of linguistically distinct tribal groups of India. *Hum Genet* 109:339–350
- Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15:174–175
- Saha A, Sharma S, Bhat A, Pandit A, Bamezai R (2005) Genetic affinity among five different population groups in India reflecting a Y-chromosome gene flow. *J Hum Genet* 50:49–51
- Schneider S, Roessli D, Excoffier L (2000) Arlequin version 2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland
- Seo Y, Uchiyama T, Matsuda H, Shimizu K, Takami Y, Nakayama T, Takahama K (2002) Mitochondrial DNA and STR typing of matter adhering to an earphone. *J Forensic Sci* 47:605–608
- Sherry ST, Rogers AR, Harpending H, Soodyall H, Jenkins T, Stoneking M (1994) Mismatch distributions of mtDNA reveal recent human population expansions. *Hum Biol* 66:761–775
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595

- Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, Singh L (2005) Reconstructing the origin of Andaman Islanders. *Science* 308:996
- Yao YG, Nie L, Harpending H, Fu YX, Yuan ZG, Zhang YP (2002) Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *Am J Phys Anthropol* 118:63–76
- Yao YG, Zhang YP (2002) Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan Province: new data and a reappraisal. *J Hum Genet* 47:311–318