Deposited research article

Molecular insight into the genesis of ranked caste populations of western India based upon polymorphisms across non-recombinant and recombinant regions in genome

Sonali Gaikwad[1] and VK Kashyap[1,2*]

Addresses: [1]National DNA Analysis Center, Central Forensic Science Laboratory, Kolkata -700014, India. [2]National Institute of Biologicals, Noida-201307, India.

Correspondence: VK Kasyap. E-mail: vkk2k@hotmail.com

Posted: 19 July 2005

Genome ***Biology*** 2005, **6:**P10

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2005/6/8/P10

Received: 18 July 2005

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

Molecular insight into the genesis of ranked caste populations of western India based upon polymorphisms across non-recombinant and recombinant regions in genome

Sonali Gaikwad [1] and VK Kashyap [1, 2*]

[1]National DNA Analysis Center, Central Forensic Science Laboratory, Kolkata -700014, India

[2]National Institute of Biologicals, Noida-201307, India

[*]Corresponding author

Email addresses:
SG: sonali_gaikwad@rediffmail.com
VKK: vkk2k@hotmail.com

Abstract

Background

Large-scale trade and cultural contacts between coastal populations of western India and Western-Eurasians paved for extensive immigration and genesis of wide spectrum of admixed gene pool. To trace admixture and genesis of caste populations of western India, we have examined polymorphisms across non-recombining 20 Y-SNPs, 20 Y-STRs, 18 mtDNA diagnostic sites, HVS-1 plus HVS-2 regions; and recombining 15 highly polymorphic autosomal STRs in four predominant caste populations- upper-ranking Desasth-brahmin and Chitpavan-brahmin; a middle-ranking Kshtriya Maratha; and a lower-rank peasant Dhangar.

Results

The generated genomic data was compared with putative parental populations- Central Asians, West Asians and Europeans using AMOVA, PC plot, and admixture estimates. Overall, disparate uniparental ancestries, and l.1% G_{ST} value for biparental markers among four studied caste populations linked well with their exchequer demographic histories. Marathi-speaking ancient Desasth-brahmin shows substantial admixture from Central Asian males but Paleolithic maternal component support their **Scytho-Dravidian** origin. Chitpavan-brahmin demonstrates younger maternal component and substantial paternal gene flow from West Asia, thus giving credence to their recent **Irano-Scythian** ancestry from Mediterranean or Turkey, which correlated well with European-looking features of this caste. This also explains their untraceable ethno-history before 1000 years, brahminization event and later amalgamation by Maratha. The widespread Palaeolithic mtDNA haplogroups in Maratha and Dhangar highlight their shared **Proto-Asian** ancestries. Maratha males harboured Anatolian-derived J2 lineage corroborating the blending of farming communities. Dhangar heterogeneity is ascribable to predominantly South-Asian males and West-Eurasian females.

Conclusions

The genomic data-sets of this study provide ample genomic evidences of diverse origins of four ranked castes and synchronization of caste stratification with asymmetrical gene flows from Indo-European migration during Upper Paleolithic, Neolithic, and later dates. However, subsequent gene flows among these castes living in geographical proximity, have diminished significant genetic differentiation as indicated by AMOVA and structure.

Background

Megadiversity in India is ascribable to diverse gene pools constituting Proto-Australoid, Caucasian Mediterranean, Mongolians and composite ethnic strains. Technological innovations outside India resulted in demic diffusion of Neolithic farmers and migration of Indo-European (IE) speakers onto Indian sub-continent [1]. Archeological and linguistic evidences support the communal social system in Indo-European tribes practicing "Andronovo culture" in Central Asia. The era 4000-1500 B.C. (Holocene period) witnessed the arrival of Indo-Aryan tribes equipped with superior military power; spread of new technologies and Indo-European language among three-quarter of contemporary Indians [2, 3]. The changed socio-cultural dimensions resulted in the cultural stratification of Indo-European and Dravidian speakers into autochthones tribes (~450 groups, 8.08%) and Hindu caste fold (80%) with many regional sub-castes [1, 4]. The caste system in India configured into four-layers was based upon professions: **Brahmin** is a priestly and learned upper-rank; **Kshatriya** is warrior and aristocratic middle-rank; **Vaishya** is a lower-ranking landowner and trading caste; and **Shudra** are social workers forming the fourth-rank [5, 6, 7]. The Caucasoid invaders: Greeks, Parthians, Scythians, and Kushans were assimilated in **Kshatriya** cluster while the pre-existing small warrior communities did not get the caste status [8, 9]. The caste endogamy allowed subsequent generations to be priests, warrior and businessmen of the society [10] and this social mechanism generated an incredibly complex genomic sub-structuring of Indian gene pool [11, 12]. Historically, northwest India first saw the permanent human settlements due to sustainable food production, storage plus cooking of food grains in pottery-ware. The trade development knitted the local social fabric into village units [13].

Present study explores the existing social mosaic (caste structure) in western India (Maharashtra's Konkan coast and north Deccan plateau) by gaining genomic insights into the

genesis of four predominant, ranked caste populations including Desasth-brahmin, Chitpavan-brahmin, Maratha, Dhangar (Table 1). The selected populations speak either "Marathi" or "Konkani"- regional languages belonging to southern branch of Indo-European family (Ethnologue Web Site). The socio-cultural environment had been greatly influenced by the en route caravan trade between north and south zone. The flourishing trade contacts with the technologically advance Western-Eurasian cultures such as Greece, Rome, and Iran resulted in exchange of food spices, cotton, leather-merchandise, and peacock-feathers during1500 B.C. - 1200 A.D. period [14] and paved the way for immigration of white sailors/merchants in different time zones and growth of few Indo-European fusion groups. Bene-Israelis (100-525 A.D), Parsis (751 A.D), Anglo-Indians (15^{th} century) and Indo-Portuguese (14^{th} - 15^{th} century) communities are settled in western coastal areas.

Previous studies based on low-resolution markers provided preliminary direction into the understanding of gene pool of western populations. Risley in 1915 [15] conducted the first ever study on Maharashtrian Brahmins using 9 anthropometric markers and speculated their "Scytho-Dravidian" origin. Later, Karve and Malhotra (1968) [16] implied different origins of Maharashtrian castes on the basis of classical markers. Baig et al. (2004) [17] used mitochondrial DNA polymorphisms to indicate early late Pleistocene maternal roots for both tribal and few caste populations of Maharashtra. However, the exchequer and varied migration histories of selected four populations confer them unique and significant from a genetic perspective and justify in-depth and extensive genome analyses. Taking these assumptions, we performed comprehensive analyses on battery of sensitive DNA markers including non-recombining paternally transmitted Y-chromosome and maternally transmitted mtDNA; and recombining biparental autosomal STR loci in four selected castes. Innovative and precise screening techniques allowed rapid population analyses for autosomal genetic markers [6, 18]; mtDNA haplotypes and haplogroup affiliations [19]; and Y-haplotypes on

stable paternal lineages identified by Y-SNPs [20]. The evolutionary histories of different haplogroups, their inferred origin and expansion through the world provide basis for reconstructing and dating prehistoric and historic population movements and scrutinize gender-specific maternal [19, 21] and paternal [22] gene flow. The higher female migration rate due to patrilocality explained greater inter-population differences at Y-chromosomes than mtDNA [23]. The three sets of DNA markers have provided genetic evidences on the origins of few of the Indian populations [11, 19, 24, 25].

The study deals with extensive comparisons of generated genomic data on recombinant and non-recombinant regions in gene pool of studied castes [17, 26, 27] with putative West-Eurasian parental populations including Central Asians [21, 28, 29]; West Asians [19, 30-34]; Europeans [22, 35 - 42], who were historically known to have large-scale business, socio-cultural and genetic interactions with populations of northwest India (see Table 1 for relevant information). These analyses address the issues surrounding genetic structure of western caste groups and resolved two major hypothesis: 1) genesis (ancestries) of Chitpavan-brahmin and Desasth-brahmin, both known to have different ethno-histories; 2) genesis of Maratha- a warrior group from agrarian, heterogeneous peasant (Dhangar) group.

Results

Y-chromosomal polymorphisms

The frequency distribution of 10 Y-haplogroups: H, R1a1, R1a, R2, L, J2, C, K2, P* and F* in 121 caste samples including published data on Chitpavan-brahmin is set in table 2. Overall frequency ranged from 0.4% (F*) to 23% (H). H and R1a1 are the most common lineages with >20% occurrences. Haplogroup H shows twofold higher and comparable frequency in Maratha and Dhangar compared to brahmin castes. Clades R2, J2, L and R1a showed appreciable frequencies above 10%. The sister clades of R1 showed clinal pattern, where R1a1 was strongly represented in Desasth-brahmin (37%) and R1a had very high frequency

(32%) in Chitpavan-brahmin from "Konkan" as compared to <=5% in other castes. Indian-specific R2 and L accounts for 29.4% Dhangar and 17% Chitpavan-brahmin chromosomes respectively. Anatolian-derived J2 lineage occurred two-times more in Maratha than other castes. Haplogroup C associated with early coastal migration was present at ≤ 5% frequency in four castes. Mean Y- haplogroup diversity was high (0.81 to 0.84) in four caste populations.

A Y-haplotype consisted of 20 Y-STR loci. Table 3 (see additional file 1) shows the listing of 77 distinct Y-haplotype configurations observed in 78 "Marathi"- and "Konkani"-speakers. Haplotype diversities were high (1.000) and similar in *"Marathi"* castes whereas Chitpavan-brahmin shared one haplotype on L background. Haplotypes were not shared between castes. Considering Y-haplogroups with frequency > 10%, mean STR variance was estimated to be highest in L (1.464), followed by J2 (1.383), R1a (1.335), H (1.076), R2 (1.002) and R1a1 (0.997). Among 20 Y-STRs, DYS426 showed 1 or 2 alleles across all clades. The modal allele (most frequent) at DYS48 was 9 on H background compared to 11 on R* and C chromosomes. The analyzed samples were two-three step mutations away from "Cohen Model Haplotype" of Near East Jews.

Mitochondrial DNA diversity

Table 4 (see additional file 2) represents 75 different mtDNA haplotypes (HVS-1 and HVS-2) in 77 individuals as one haplotype each was shared within Maratha and Chitpavan-brahmin respectively. The HVS-1 sequence motifs and associated diagnostic mutations clustered into 25 clades and sub-clades belonging to macrohaplogroups M, N, and R. Superhaplogroup M is partitioned into 9 sub-haplogroups; shows highest frequency (64%) followed by R (18.7%), U (14.6%) and minor fraction of N (2.8%). Maharashtrain R clade include R*, H, HV, U*, K, J2; and N comprise of N*, W. Phylogenetic relationships among HVS-1 haplotypes falling into different haplogroups are presented (Figure 1).

The Early Upper Paleolithic South-Asian cluster M has the highest frequency followed by M5, M4, M8c, and M25 etc. M lineages were most frequent (72%) in Maratha followed by Chitpavan-brahmin (67%), Desasth-brahmin (53%) and Dhangar (47%). The other South-Asian specific diverse sub-clusters of R and U were also frequent in Maratha (89%), and Chitpavan-brahmin (84%). U7 and W lineages associated with another late upper Paleolithic migration to Indian sub-continent accounted for 1% (CB) to 10% (DB). Neolithic migration and very recent Western influence was seen more (14-15%) in Chitpavan-brahmin and Desasth-brahmin than the other two castes. The studied castes did not carry A, B, M-C and M-D haplogroups as indicated by - 663 *Hae*III, + 5176 *Alu*I, + 13259 *Hinc*II sites and absence of 9 bp deletion between COII / $tRNA^{Lys}$.

A set of mtDNA diversity indices and time of demographic expansion for four castes is shown in Table 5. Haplotype diversities were highest (1.000 ± 0.017) in Dhangar and Desasth-brahmin while haplotypes were shared within Maratha and Chitpavan-brahmin. The studied groups did not share haplotypes among them. The nucleotide diversities and mean pair-wise differences were found to be higher and similar in Maratha and Dhangar than the two brahmin castes. The four castes showed unimodal (bell-shaped) mismatch distributions (figure not shown). In parallel, the raggedness index was less than 0.05, negative values of Fu's Fs and Tajima's D differ significantly from zero. These values agreed well with mismatch analysis and provided clear evidence of demographic expansion of castes from western India.

Tau value based upon the location of the mismatch distribution crest provides a rough time estimate of rapid expansion of population [43]. The estimated values were larger than 6.0 corresponding to expansion times of >52,000 YBP while Chitpavan-brahmin show a small value of 4.59 conforming to 39,000 YBP.

Autosomal STR analyses

All studied populations exhibited no detectable deviation from H-W Expectation with 15 loci. The segregating alleles were found to be higher and similar (159) for Maratha and Dhangar, intermediate (153) in Desasth-brahmin and lower (142) in Chitpavan-brahmin. DH and CB shared most frequent alleles at 9 loci and rare alleles ranging from 0.6 - 0.7% at 4 loci [26].

Single locus heterozygosity (Ho) estimates ranged from 0.612 (CSF1PO; CB) to 0.948 (Penta E; DH). Combined heterozygosity was remarkably high and comparable in four castes, ranging from 77% (MA) to 80% (DB). Average variance in allele size considering 15 loci was estimated to be slightly higher (4.021) in Chitpavan-brahmin, followed by Dhangar (3.969), Maratha (3.920) and Desasth-brahmin (3.732).

Genomic variations between Maharashtrian castes

The G_{ST} estimate at biparental markers varied widely from 0.002 (D18S51) to 0.041 (D7S820); combined value for all the loci considered together was 0.011 indicating low genetic differentiation between populations. Using structure program, the proportions of individuals assigned to each cluster were approximately the same with little variation between ethnic groups under the admixture model. This symmetry is strongly suggestive of the absence of population structure in the present study, since real population structure is associated with individuals being strongly assigned to one inferred cluster or another with the proportions assigned to each ethnic group showing asymmetry (Figure 2)

Population clustering as revealed by PCA

Genetic affinities between studied castes and West Eurasians (see Table 1) were tested in the light of population histories using Principal Component analyses. Results of PCA are presented by plots of the first two PCs, which together account for 64% of the Y-

chromosome haplogroup, 68% of mtDNA variation and 77% of the biparental variations in these populations.

Figure (3a) represents the PCA of Y-haplogroups, where Maratha, Desasth-brahmin and Dhangar cluster together while Chitpavan-brahmin appears as an outlier, which show genetic affinities with West Asians (Ashkenazi-Jews, Iranian); Greeks and Central Asians compared to very distinct West-European French and Portuguese. PC plot with basal mtDNA haplogroup frequencies as input vectors is shown (Figure 3b), where the first two PC account for 42.165% and 25.724% of the total variation respectively. The first PC mainly separates Desasth-brahmin from Chitpavan-brahmin, the latter caste occupied intermediate position between Indian and West Eurasian (Uzbeks, Turks, Iranian). The populations from West Asia and Europe tend to cluster together. A PC plot using F_{ST} distance matrix based on 7 autosomal STR loci (Figure 3c) showed 77% variance accounting for first two components indicating a satisfactory representation of the original data. Dhangar (DH) appeared as an outlier. Chitpavan-brahmin (CB), Desasth-brahmin (DB) and Maratha (MA) cluster together and were placed closer to the West-Eurasian populations.

Evaluation of hierarchical structure, using AMOVA

Recombining and non-recombining DNA polymorphisms showed >90% variations within populations (Table 6). The studied castes showed maximum variance (5.29%) at Y-hapolgroups followed by mtDNA sequence diversity. Considering socio-hierarchy, the upper, middle, and lower-ranking castes showed insignificant variance while statistically significant variance was apportioned between two Brahmin castes based on their Y-SNP and mtDNA data. Maratha and Dhangar showed greater variance at Y-haplogroup followed by mtDNA. However, “Marathi” and “Konkani” Indo-European language groups reveal insignificant genetic variance, irrespective of genomic dataset. A moderate difference was observed between the mtDNA sequence diversity of Marathi-speakers. Considering

AMOVA analyses using haplogroup frequency data of studied castes, Central Asian, West Asian and European populations, Y- haplogroup showed maximum genetic variance among analyzed populations (9.14%) compared to mtDNA lineages (4.36%). The studied castes irrespective of rank affiliation showed lesser genetic variance with Central Asians and West Asians compared to Europeans.

Western Eurasians admixture in western castes

Admixture estimates based on three sets of genetic markers are quantified and summarized (Table 7). All four castes show very high (60-90%) indigenous genetic component. The Western Eurasian admixture ranged from 10 - 40%. Male-specific admixture from Mediterranean belt was highest (40 – 50%) in Chitpavan-brahmin; from Central Asia and East-Europe was 22 – 28% in Desasth-brahmin; and from Central Asia and West Asian was 20-25% in Maratha. Maternal gene flow from West Asians was prominent in Dhangar and Desasth-brahmin; European admixture was moderate in Chitpavan-brahmin.

Discussion

Comprehensive analyses of recombining biparental and non-recombining uniparental markers helped in resolving the scientific lacunae surrounding genetic structure, affinities, and origin of ranked castes of western India. Earlier studies on western Indian populations were based on low-discriminating genetic markers, which provided interesting preliminary information on their genetic structure and origins [16, 17, 44]. However, the composite population history could not be deciphered because it lacks the very essential Y-chromosomal component.

Present study gained deep and fresh insights into the maternal and paternal ancestry and biparental composition of four hierarchical caste populations practicing diverse socio-cultural traditions to test the generality of various competing hypothesis on the origins of Indian caste populations proposed in aforementioned studies. Convincing genomic evidences

for the diverse origins of four western caste populations was substantiated by the presence of finer branches of major mtDNA and Y-chromosome haplogroups dating back to early Upper Paleolithic (30000-50000 YBP) or recent coalescence age (~10, 000 years ago). The higher proportion of mtDNA lineages (U2, W, R5, U7) in four castes suggested ancestral gene pool. However, West-Eurasian particularly Central Asian and East European-specific male lineages (R1a, R1a1, J2), and mtDNA lineages (H, K, HV, J2, U5, U3) were present at variable frequencies in studied caste groups strongly indicating admixture and ancestry with latest migrants. Autosomal microsatellite diversity show high heterogeneity (78.4%) and large number (622) of segregating alleles across 15 loci; however the three social ranks did not reveal significant genetic variance and low coefficient of gene differentiation (G_{ST} = 1.1%) among them indicated genetic affinities among western populations. The above result also implies recent common origin of some groups and substantial gene flow among the similar ranking castes.

The genetic affinity with Western-Eurasians is explained in the light of immigration of trade merchants to "Konkan" west coast and conquest of Scythians/ Sakas, Kushans and White-Huns over Indo-Greek in Indus Valley during 1st to 5th century B.C. [45, 46]; subjugation of natives in northwest India, formation of "Indo-Scythian" fusion groups, notably the "Rajput", the "Mauryans" and "western Ksatrapas" in western India, who overthrew "Satavahans" dynasty in "Deccan" plateau (Maharashtra) to gain control over the caravan trade routes and also supported Buddhist cave monasteries by the generated wealth.

Genesis of Brahmin castes

Chitpavan-brahmin and Desasth-brahmin constitutes just 10% of entire populace (~80 million) of western India. Their different marriage rules, varied customs, different local dialects (southern branch of Indo-Aryan language) illustrated distinct origins. Our extensive comparative analyses support their different ethno-histories. The Y-chromosomes of Marathi-

speaking Desasth-brahmin carried R1a1 lineage in high frequency, which reflected their considerable affinity with Central Asian giving credence to their "Scythic" descent (admixture, PC plot, AMOVA analysis). Their intermediate mtDNA diversity comprise of low frequency West-Eurasian clades and significant Paleolithic gene pool (M) indicating South-Asian ancestry, which provide evidence of their tribal origin due to upward social mobility of females as shown by study of Baig et al. (2004). These Brahmin subjects presented highest number of biparental alleles, heterozygosity and genetic affinity with Central Asians. These analyses provide evidence of "**Scytho-Dravidian**" genesis of Desasth-brahmin. They are the ancient upper-caste comprising of 50 sub-divisions or "gotra" [47] because the considerable time-depth as inferred from Tau value, helped them to consolidate their predominance in different administrative jobs besides traditional priesthood.

Conversely, non-recombining uniparental contributions in Chitpavan-brahmin Mediterranean or East European type as shown by 20% (HV, U3) mtDNA lineages and highly frequent (R1a and L) Y-haplogroups. The admixture and PC analyses (Figure 3a, b) reflected genetic association of Chitpavan-brahmin with Iranian, Ashkenazi-Jews (Turkey), Greeks (East Europe) and to some extent with Central Asian Turkish populations elucidating their distinct Nordic, "**Scytho-Iranian**" ancestry [48, 49]. The Caucasian link of Chitpavan-brahmin has also been inferred from biparental microsatellites variations (Figure 3c). The observed genomic analyses asserted the ethnographical fact that Chitpavan-brahmin share ancestry with conspicuously European-looking Pagan or Alpine group, who under religious pressure had migrated from Anatolian Turkey or East Europe to Gujarat coast probably via sea-vessel. Besides, their documented history is untraceable beyond 1000 years, further indicating that they were not part of the original Vedic migrations (early Indo-European) on the west coast. Therefore, the present genome analyses provide conclusive evidence of their recent migration, genesis, and expansion after they migrated from "Sopara" (India's western

trade zone) to geographically isolated Konkan-region, where they adopted "Konkani" language, and cultivated cash crop. Their considerable genetic affinity with Maratha caste further corroborated the prevalent norm that few of the dynamic and intelligent Chitpavans were "Brahmanized" for performing religious rituals in King Shivaji's court (elite Maratha group) and some members were given the title of "Peshwa" or Minister for managing the administration of Maratha kingdom, which was extended farther north after King's death under their rule. We observed 15% similar HVS -1 sequence motif (M4 lineage) between Chitpavan-brahmin and Bene-Israeli (or Indian Jews), probably suggesting similar indigenous Paleolithic contribution. Compared to Desasth-brahmin, Kokanasth-brahmin showed lowest biparental diversity, younger age of population based upon Tau value, larger genetic affinity with West Asians plus East Europeans suggesting their recent descent, in absence of bottleneck effect. However, recent marriages between Desasth-brahmin boy and Chitpavan girl have contributed towards their genetic affinity as shown structure plot (Figure 2).

Origin of peasantry Dhangar caste

The 23 endogamous sub-castes of pastoral caste cluster show considerable variation in their numerical strength and ecological distributions. The complex ethno-history suggests expansion and contraction during the long evolutionary history of this peasant cluster. Our analyses on Y-haplogroups clearly elucidated Proto-Asian genetic ancestry of Dhangar, whose Y- haplogroup diversity was slightly lower than Maratha evidently supporting fission in Dhangar cluster and fusion in Maratha. Contrastingly, high levels of West Asian maternal component in Dhangar suggested asymmetrical gene flow. The 15 biparental microsatellite markers showed rich allelic diversity and high heterozygosity owing to cluster expansion. The process of genetic fission and fusion of linguistically (Marathi, Hindi, Kannada, Telugu) and occupationally diverse groups is better understood by high levels of diversity in Dhangar's sub-castes of southern and northeastern Maharashtra [50]. An earlier study based

on bilateral palmar prints in 20 sub-castes [51] provide explanatory evidences on fission of single caste into sub-castes, namely, Hatkar, Zende, Thellari, and Dange; and fusion of linguistically different subgroups such as Ahirs, Shegars from northwest India and Kurmars from south India for carrying out similar occupational pursuits. The practice of inbreeding occurs in some isolated groups formed as a result of fission [51]. Thus, the existing sub-structuring within a Dhangar peasant caste is attributable to innumerable fission and fusion processes for demographic and economic reasons.

Genesis of warrior Maratha caste

Maratha caste is geographically dispersed and represents more than 50% of the current Maharashtrian population. This endogamous community of fifty- million individuals became dominant owing to their occupational hierarchy. Their warrior element is conglomeration of Royal descendents such as Rashtrakuts, Mauryas, Pariharas/ Parmar (Pawar), Pratiharas, Shilahars, Kadambas, Yadavas, Chalukyas etc. as a result of successful expeditions and conquests of different parts (small Kingdoms) of the country. Our analyses showed limited frequency of Holocene-specific mtDNA (U5, H and W) but higher frequency of South-Asian lineages, substantiating their Paleolithic ancestry. Maratha shows higher nucleotide diversity, mean-pairwise differences than Brahmin castes but comparable with Dhangar suggesting common origins as a result of gene flow from peasant cluster as shown by structure plot (Figure 2). It also depicted affinity with two Brahmin caste, which clearly more than expected matrimonial alliances between upper and middle-ranking castes [6, 7]. Maratha males, in relation to studied castes, carried higher frequency of J2 lineage [22], which strongly indicated assimilation of Anatolian farmers into Maratha gene pool. The above observation once again asserted the fact that Kshtriya group has mixed gene pool because of its fluid genetic boundary [3, 45].

Conclusions

Our comprehensive genomic analyses showed divergent paternal and maternal ancestry of studied four castes correlating well with their varied migration and exchequer demographic histories. The distribution and admixture of Western-Eurasian-specific mtDNA and Y-chromosomal haplogroups lend support to the diverse genesis of western ranked castes. The asymmetrical Proto-Asian component and Western-Eurasian admixture in two brahmin castes explained the "Scytho-Dravidian" origin of elite, ancient Desasth-brahmin and much recent "Irano-Scythian" ancestry (West Asia, East Europe) of Chitpavan-brahmin. Maratha and Dhangar have significant Pleistocene gene pool corroborating their "Proto-Asian" origin. Maratha warrior caste has experienced gene flow from Anatolian agriculturist (J2) supporting the conglomeration of migrant agricultural communities. The recombining STR loci did not reveal significant difference in population structure attributing to hypergamy between Brahmins and Maratha, and shared ancestry of Dhangar and Maratha. This study interestingly surmises the synchronization of caste stratification with West-Eurasians admixture in "Gangetic" plains, which spread in western territory due to demographic and economic reasons.

Methods

The Populations

In this study, blood specimens were collected in K_3EDTA vials from 365 unrelated, healthy and consenting individuals belonging to Desasth-brahmin, Chitpavan-brahmin, Maratha, and Dhangar castes inhabiting Maharashtra (20^o N latitude, 76^o E longitude), western India (see Maps of India web site). The study has been undertaken with the approval of Ethical Committee of CFSL (Kolkata) and MHA, Government of India.

The sampling locations, relevant demographic information, and data sources regarding caste and Western-Eurasian populations analyzed for 3 sets of markers are given in Table 1. Available Y- and mtDNA haplogroup data of Chitpavan-brahmin and Maratha (see table 1 for specific citations) are included in our dataset. The Western-Eurasian populations included for comparative analyses were Central Asians [Uzbek, Turkish, Kurd); West Asians [Ashkenazi-Jews, Arabs, Iranian]; and European [Portuguese, Greeks, French] (see Table 1 for references). Due to paucity of published autosomal STR datasets, only 7 loci [vWA, TH01, D18S51, D3S1358, D21S11, D8S1179, FGA] in Turkish, Ashkenazi-Jews, Iranian, and three European populations were considered to explore genetic affinities and gene flow patterns.

Sequencing and Genotyping study

Genomic DNA was isolated by standard phenol-chloroform method [52].

Y-chromosome typing at 20 bi-allelic and 20 multi-allelic markers was performed in 78 males out of 120 sampled males from 4 castes. Y-SNPs: M168, YAP, RPS4Y, M122, M89, M172, M69, M9, 92R7, M3, M207, M173, SRY1532, M17, M124, M18, M5, M20, M11, and M70 known to identify 12 haplogroups in Eurasian populations were typed using validated amplification protocols [20, 53, 54]. Y- haplogroup nomenclature was done following Y-Chromosome Consortium [55].

The Y-haplogroups with related haplotypes are distinguishable via microsatellite markers. A single multiplex fluorescent-based genotyping assay of 20 Y-STRs (Table 3, see additional file 1) was done following primer and amplification conditions [56]. STR amplicons of two dinucleotide, 16 tetranucleotides, and two pentanucleotides (table 3 for loci names) were separated on ABI 3100 Genetic Analyzer, sized by Genescan^{TM} 3.1 and assigned allele number for complete Y-haplotype profile.

Seventy-seven males representing 4 castes were sequenced at HVS-1 (15997-16391) and HVS-2 (48-408) fragments using Big-Dye^{TM} terminator chemistry, purified by ethanol precipitation and resolved on ABI 3100 Genetic Analyzer (Applied Biosystems). Super-haplogroups were defined based on RFLP diagnostic markers [57]. The sub-haplogroup or finer lineages were assigned by assaying additional informative sites (Table 4, additional file 2) by either complete mtDNA sequencing [58] or diagnostic markers in coding-segments [19, 21, 59]. Both HVS-1 motif and coding-region variations were used to classify the maternal lineages according to the above mentioned published sources.

All 356 sampled individuals were genotyped at 15 STR loci [two pentanucleotides and 13 tetranucleotides] co-amplified using PowerPlex® 16 multiplex kit and manufacturer's instructions (Promega Corp., Madison, USA).

Data Analyses

Y-STR haplotypes were constructed for caste samples (present study). Y-STR variance was estimated for 6 common haplogroups. mtDNA mutations were scored with reference to revised Cambridge Reference Sequence [60]. Reduced-Median Network was drawn for HVS-1 haplotypes and their diagnostic markers in caste subjects (A Rohl; Shareware Phylogenetic Network Software Web Site). Fifteen biparental STR analyses include estimation of heterozygosity [61], allele size variance [62] and coefficient of gene differentiation/ G_{ST} [63] based on published allele frequencies in four caste populations [26]. The diversity indices and demographic parameters, viz., Tau value, Tajima's D, and Fu's F_S tests were estimated using Arlequin software, version 2.001 [64].

To substantiate the hypothesis of common ancestry of Maratha and Dhangar, and gene flow between two Brahmin castes, we analyzed genotype data of unlinked markers via admixture-model of Structure software, version 2.0 [65]. Each run was done after 100,000 burn-in iterations and 1,000,000 estimation iterations for K=1 to 5

The genetic relationships between studied castes and historically known putative West-Eurasian parental populations (Table 1) were examined via Principal-component analysis (PCA) using SPSS 11.0 package. The input variables were uniparental haplogroup frequencies and F_{ST} distance matrix for 7 biparental STR loci. AMOVA approach was used to estimate proportion of genetic variance in caste and reference populations using ARLEQUIN package. Detailed grouping designs are listed in Table 6.

ADMIX95 Software based on Gene Identity method [66] was used to estimate the admixture proportions (m $\pm$ SE) of West-Eurasian populations in the western caste populations. The putative parental populations were known to have historical trade and cultural links with India.

Authors' contributions

SG performed all laboratory experiments, carried out statistical analyses and drafted the manuscript. VKK conceptualized the hypothesis of the study, helped in correct interpretation of results and improved the presentation and style of the manuscript.

Acknowledgements

Authors express their heartfelt gratitude to all blood donors from western India; Director and Dr. R. Trivedi, Chief, National DNA Analysis Center, Central Forensic Science Laboratory, Kolkata for providing laboratory facilities and support in analyses. The paper would not have been possible without extensive information on ethno-historical background on studied caste populations and intellectual inputs of Dr. TS Vasulu, Indian Statistical Institute, Kolkata. The first author acknowledges student scholarship from DFS, Ministry of Home Affairs, Govt. of India.

References

1. Gadgil M, Shambu Prasad UV, Manoharan S, Patil S, Joshi NV: **Peopling of India**. In: *The Indian human heritage*. Edited by: Balasubramanian D, Appaji Rao N. Hyderabad: Universities Press; 1997:100-129.
2. Renfrew C: *Archaeology and language: the puzzle of Indo-European origins*. London: Jonathan Cape; 1987.
3. Quintana-Murci L, Krausz C, Zerjal T, Sayar SH, Hammer MF, Mehdi SQ, Ayub Q, Qamar R, Mohyuddin A, Radhakrishna U, et al.: **Y-chromosomes lineages trace diffusion of people and languages in southwestern Asia**. *Am J Hum Genet* 2001, **68**:537-542.
4. Majumder PP: **People of India: Biological diversity and affinities**. *Evol Anthropol* 1998, **6:**100-110.
5. Karve I: *Kinship organization in India.* Bombay: Asia Publishing House; 1968.
6. Cavalli-Sforza L, Menozzi P, Piazza A: *The history and geography of human genes*. Princeton: Princeton University Press; 1994.
7. Singh KS, ed: *People of India: India's communities*. Volumes IV-V. New Delhi: Oxford University Press; 1998.
8. Spear P: *A history of India*. Volume 2. London: Penguin Books; 1975.
9. Sastri KAN, Srinivasachari G: *Advanced History of India*. 2nd edition. New Delhi: Allied Publishers Pvt. Ltd; 1980.
10. Majumder PP: **Indian caste origins: Genomic insights and future outlook**. *Genome Res* 2001,.**11**:931-932.
11. Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BV, Reddy PG, Rasanayagam A, et al.: **Genetic evidence**

on the origins of Indian caste populations. *Genome Res* 2001, **11**:994-1004.

12. Roychoudhury S, Roy S, Basu A, Banerjee R, Vishwanathan H, Usha Rani MV, Sil SK, Mitra M, Majumder PP: **Genomic structures and population histories of linguistically distinct tribal groups of India**. *Hum Genet* 2001, **109**:339-350.

13. Bhasin MK, Walter H: *Genetics of castes and tribes of India*. New Delhi: Kamla-Raj Enterprises; 2001.

14. Klostermaier KK: *A survey of Hinduism*. 2nd edition. New York, Albany: SUNY Press; 1994.

15. Risley HH: *The people of India*. Delhi: Orient Book Reprint Corporation; 1915.

16. Karve I, Malhotra KC: **A biological comparison of eight endogamous groups of the same rank**. *Curr Anthropol* 1968, **9**:109-124.

17. Baig MM, Khan AA, Kulkarni KM: **Mitochondrial DNA diversity in tribal and caste groups of Maharashtra (India) and its implication on their genetic origins.** *Ann Hum Genet* 2004, **68**:453-460.

18. Zhivotovsky LA, Rosenberg NA, Feldman MW: **Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers.** *Am J Hum Genet* 2003, **72**:1171-1186.

19. Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Thomas M, et al.: **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* 2004, **5**:26.

20. Underhill PA, Passarino G, Lin AA, Shen P, Mirzazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL: **The phylogeography of Y-chromosome**

binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 2001, **65**:43-62.

21. Quintana-Murci L, Chaix R, Spencer Wells R, Behar DM, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti AS, et al.: **Where West meets East: the complex mtDNA landscape of the southwest and central Asian corridor**. *Am J Hum Genet* 2004, **74**:827-845.

22. Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, et al.,: **Y-chromosomal diversity in Europe is clinal and is influenced primarily by geography, rather than by language**. *Am J Hum Genet* 2000, **67**:1526-1543.

23. Seielstad MT, Minch E, Cavalli-Sforza LL: **Genetic evidences for a higher female migration rate in humans**. *Nat Genet* 1998, **20**:278-280.

24. Cordaux R, Saha N, Bentley GR, Aunger R, Sirajuddin SM, Stoneking M: **Mitochondrial DNA analysis reveals diverse histories of tribal populations from India**. *Eur J Hum Genet* 2003, **11**:253-264.

25. Cordaux R, Aunger R, Bentley GR, Nasidze I, Sirajuddin SM, Stoneking M: **Independent origins of Indian caste and tribal paternal lineages**. *Curr Biol* 2004, **14**:231-235.

26. Gaikwad S, Kashyap VK: **Polymorphism at fifteen hypervariable microsatellite loci in four populations of Maharashtra, India**. *Forensic Sci Int* 2002, **126(3)**: 267-271.

27. Kivisild T, Rootsi M, Metspalu S, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk H-V, Stepanov V, et al.: **The genetic heritage of earliest settlers persists both in Indian tribal and caste populations**. *Am J Hum Genet* 2003, **72**: 313-332.

28. Akbasak BS, Budowle B, Reeder DJ, Redman J, Kline MC: **Turkish population data with the CODIS multiplex short tandem repeat loci**. *Forensic Sci Int* 2001, **123**:227-229.

29. Zerjal T, Spencer Wells R, Yuldasheva N, Ruzibakiev R, Tyler-Smith C: **A genetic landscape reshaped by recent events: Y-Chromosomal insights into Central Asia**. *Am J Hum Genet* 2002, **71**:466–482.

30. Picornell A, Tomas C, Jimenez G, Castro JA, Ramon MM: **Jewish population genetic data in 20 polymorphic loci**. *Forensic Sci Int* 2002, **125**:52-58.

31. Nebel A, Filon D, Brinkmann B, Majumder PP, Faerman M, Oppenheim A: **The Y chromosome pool of Jews as part of the genetic landscape of the Middle East**. *Am J Hum Genet* 2001, **69**:1095–1112.

32. Behar DM, Hammer MF, Garrigan D, Villems R, Bonne-Tamir B, Richards M, Gurwitz D, Rosengarten D, Kaplan M, Pergola SD, et al.: **MtDNA evidence for a genetic bottleneck in the early history of the Ashkenazi Jewish Population**. *Eur J Hum Genet* 2004, **12(5)**: 355-364.

33. Al-Zahery N, Semino O, Benuzzi G, Magri C, Passarino G, Torroni A, Santachiara-Benerecetti AS: **Y-chromosome and mtDNA polymorphisms in Iraq, crossroad of the early human dispersal and of post-Neolithic migrations**. *Mol Phylogenet Evol* 2003, **28**: 458–472.

34. Kivisild T, Rootsi S, Metspalu M, Metspalu E, Parik J, Kaldma K, Usanga E, Mastana S, Papiha S, Villems R: **The Genetics of Language and Farming Spread in India**. In: *Examining the farming/language dispersal hypothesis.* Edited by: Bellwood P, Renfrew C. Cambridge: The McDonald Institute for Archaeological Research; 2003: 215-222.

35. Santos SMM, Budowle B, Smerick JB, Keys KM, Moretti TR: **Portuguese**

population data on the six short tandem repeat loci – CSF1PO, TPOX, TH01, D3S1358, VWA and FGA. *Forensic Sci Int* 1996, **83**: 229-235.

36. Anjos MJ, Carvalho M, Andrade L, Corte-Real F, Vieira DN, Vide MC: **Allele frequencies of STR multiplex systems in two Portuguese population samples**. *Progr Forensic Genet* 2000, **8**:208-211.

37. Pereira L, Prata MJ, Amorim A: **Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation**. *Ann Hum Genet* 2000, **64**: 491-506.

38. Sanchez-Diz P, Lareu MV, Brion M, Skitsa I, Carracedo A: **STR data for the AmpFlSTR Profiler Plus loci from Greece**. *Forensic Sci Int* 2002, **126**: 265-266.

39. Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, et al.: **Tracing European Founder Lineages in the Near Eastern mtDNA pool**. *Am J Hum Genet* 2000, **67**:1251-1276.

40. Tofanelli S, Taglioli L, Varesi L, Paoli G: **The STR-based genetic profile of the population from Corsica island (France)**. *Forensic Sci Int* 2001, **123**: 33-38.

41. Rousselet F, Pfitzinger H, Mangin P: **Multiplex amplification and automated fluorescent typing of short tandem repeat (STR) loci: The French Experience**. *Adv Forensic Haemogenet* 1996, **6**:139-141.

42. Rousselet F, Pfitzinger H, Mangin P: **A pentaplex automated fluorescent typing system for forensic idenfication and French Caucasian population data**. *J Forensic Sci* 1997, **42**:500-503.

43. Rogers AR, Harpending H: **Population growth makes waves in the distribution of pairwise genetic differences**. *Mol Biol Evol* 1992, **9**:552-569.

44. Roy S, Thakur C, Majumder PP: **Mitochondrial DNA variation in ranked caste**

groups of Maharashtra (India) and its implication on genetic relationship and origins. *Ann Hum Biol* 2003, **30(4):** 443-454.

45. Thapar R: *A History of India*. Volume I. London: Penguin Books; 1975.

46. Kochar R: *The Vedic people*. Hyderabad: Orient Longman; 2000.

47. Saoji AV, ed: *Samasth Maharashtriya Brahmin Potshakha* (*Root of all Maharashtrian Brahmins,* in "Marathi"). Nagpur: Sukhad Prakashan; 2000.

48. Guha B: **The racial affinities of the people of India**. In: *Census of India, 1931, Part III Ethnographical* Simla: Government of India Press; 1935.

49. Dixit J, Series Ed: *Chitpavanism, A tribute to Chitpavan Brahmin culture, Encyclopedia Maharashtra***.** Volume 1. Pune: Dixit Publishers; 2003.

50. Malhotra KC: **Inbreeding among Dhangar castes of Maharashtra, India**. *J Biosoc Sci* 1979, **11**:397-409.

51. Karmakar B: **Distribution and population variation of total number of palmar triradii among 20 Dhangar castes of Maharashtra, India**. *Int. J. Anthropol* 1990, **5(3)**:255-269.

52. Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning: a laboratory manual.* Edited by: Ford N, Nolan C, Ferguson MC. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1989.

53. Whitfield LS, Sulston JE, Goodfellow PN: **Sequence variation of the human Y-chromosome**. *Nature* 1995, 378: 379-380.

54. Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ: **Detection of numerous Y-chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography**. *Genome Res* 1997, **7**:996-1005.

55. Y-Chromosome Consortium: **A nomenclature system for the tree of human**

Y-chromosomal haplogroups. *Genome Res* 2002, **12**:339-348.

56. Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, HammerMF: **A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers**. *Forensic Sci Int* 2002, **129**:10–24.

57. Roychoudhury S, Roy S, Dey B, Chakraborty M, Roy M, Roy B, Ramesh A, Prabhakaran N, Usha Rani MV, Vishwanathan H, et al.: **Fundamental genomic unity of ethnic India is revealed by analysis of mitochondrial DNA**. *Curr Sci* 2000, **79(9)**:1182-1192.

58. Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, Simionati B, Valle G, Richards M, Macaulay V, Scozzari R: **Do the four clades of the mtDNA haplogroups L2 evolve at different rates?** *Am J Hum Genet* 2001, **69(6)**:1348-1356.

59. Palanichamy MG, Sun C, Agrawal S, Bandelt HJ, Kong Q, Khan F, Wang C, Chaudhuri TK, Palla V, Zhang Y: **Phylogeny of mitochondrial DNA haplogroup N in India, based on complete sequencing: implications for the peopling of South Asia**. *Am J Hum Genet* 2004, **75(6)**:966-978.

60. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N: **Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA**. *Nat Genet* 1999, **23**:147.

61. Nei M: **Estimation of average heterozygosity and genetic distance from a small number of individuals**. *Genetics* 1978, **89**:583-590.

62. Minch E, Ruiz-Linares A, Goldstein D, Feldman M, Cavalli-Sforza LL: **MICROSAT, Version 1.0: a computer program for calculating various statistics on microsatellite allele data**. In: Stanford University, Stanford 1997.

63. Nei M: **Analysis of gene diversity in subdivided population**. *Proc Natl Acad Sci USA* 1973, **70(12)**:3321-3323.

64. Schneider S, Roessli D, Excoffier L: **Arlequin, Version 2.0: a software for population genetic data analysis**. In: University of Geneva, Geneva 2000.

65. Pritchard JK, Stephens M, Donnelly P**: Inference of population structure using multilocus genotype data**. *Genetics* 2000, **155**:945-959.

66. Chakraborty R: **Gene Identity in racial hybrids and estimation of admixture rates**. In: *Genetic microdifferentiation in man and other animals.* Edited by: Neel JV, Ahuja YR. Delhi University, Delhi: Indian Anthropological Association; 1985: 171-180.

Figures

Figure 1. **Reduced Median Network relating 77 mtDNA haplotypes in western Indian caste subjects.** Circle areas are proportional to haplotypes frequencies. Population codes are as reported in table 1. M and sub-clusters (blue circles), R and sub-clusters (pink circles), U and sub-clusters (green circles), N and sub-lineages (red circles)

Figure 2. **Assignment of samples from four western Indian caste populations to genetic Clusters inferred from the STRUCTURE analysis for K = 4.**

Figure 3a. **Principal Component plot based on Y-Haplogroups frequencies among castes of western India and 9 putative parental West Eurasian groups.** Table 1 for population abbreviation.

Figure 3b. **Principal Component plot based on mtDNA-haplogroups frequencies among castes of western India and 9 putative parental West Eurasian groups.** Table 1 for population codes.

Figure 3c. **Principal Component plot based on based on F_{ST} distance matrix constructed using biparental STR allele frequencies among castes of western India and 9 putative parental West Eurasian groups.** Population codes as in table 1.

Tables

Table 1. Demographic characteristics of studied populations, sample sizes, source of genomic data of studied and related population.

Table 2. Distribution of Y-haplogroups in caste populations of Western India.

Table 5. mtDNA diversity indices and demographic parameters in caste populations of western India.

Table 6. Analyses of Molecular Variance in caste and putative parental populations based upon uniparental and biparental markers.

Table 7. Admixture estimates based on bi- and uni-parental genetic markers in caste populations of western India

Additional materials

Additional File 1

Table 3. Excel spreadsheet. Distribution of Y-STR haplotypes in four caste populations of western India.

Additional File 2

Table 4. Excel spreadsheet. Distribution of mtDNA haplogroups (subclusters) with HVS -I and II sequence variations in four caste populations of western India.

Table 1

Demographic characteristics of studied populations, sample sizes, source of genomic data of studied and related populations

A. Castes				Genomic data sources and sample sizes (N)		
Populations [code]	**Geographical distribution[a]**	**Social status and ethnohistory**	**Linguistic family[b]**	**Autosomal STR (N)**	**Y-haplogroup (N)**	**mtDNA haplogroup (I**
Desasth-brahmin [DB]	Western India	Upper caste, Indo-Caucasoid pool	Marathi (IE)	102 [26]	19 (this study)	19 (this study)
Chitpavan-brahmin [CB]	Western India	Upper caste, Nordic built with light-color eyes	Konkani (IE)	67 [26]	66 = 23 (this study) + 43 [27]	77 = 20 (this study) + 57 [27]
Dhangar [DH]	Western India	Lower peasantry caste with 23 sub-clusters; genetic fusion or fission common	Marathi (IE)	80 [26]	17 (this study)	19 (this study)
Maratha [MA]	Western India	Middle caste for defense pursuits, blend of guerilla groups and agrarian classes	Marathi (IE)	107 [26]	19 (this study)	29 = 19 (this study) + 10 [17]
B. Western-Eurasians						
Uzbek [UZ]	Central Asia	agriculturalism	Altaic	Not avaliable	[29]	[21]
Turkish [TK]	Central Asia	pastoral nomadism	Altaic Turkic	[28]	[29]	[21]
Kurd [KT]	Central Asia	agriculturalism	IE	Not avaliable	[29]	[21]
Ashkenazi-Jews [AJ]	West Asia, Israel	Middle eastern ancestry, genetic bottleneck by mDNA study	Levantine Arabic	[30]	[31]	[32]
Arabs [AR]	West Asian, Saudi Arabia	ancestry with nomadic Semitic tribes, founder of Islam, seafaring trade in oil sector	Gulf Arabic	Not avaliable	[31]	[33]
Iranian [IR]	West Asia, Iran	Aryan origin from Indo-Iranian tribes, founder-Parsi religion	Farsi (IE)	[30]	[34]	[19]
Portuguese [PT]	West Europe, Portugal	technologically advance business community	Italic (IE)	[35, 36]	[22]	[37]
Greeks [GK]	East Europe, Greece	Alexander conquered Eurasia, Indo-Greek groups in northwest India	Greek (IE)	[38]	[22]	[39]
French [FR]	West Europe, France	technologically advance business community	Italic (IE)	[40, 41, 42]	[22]	[39]

[a] Maharashtra map (www.mapsofindia.com); [b] IE: Indo-European, (www.ethnologue.com)

Table 2. Distribution of Y-haplogroups in caste populations of Western India.

Y-Haplogroup	Desasth-brahmin	Chitpavan-brahmin	Maratha	Dhangar	Average Frequency
	N=19	N=66	N=19	N=17	
H	0.158	0.137	**0.315**	0.294	**0.226**
R1a1	**0.368**	0.045	0.211	0.235	**0.215**
R2	0.105	0.106	0.053	**0.294**	**0.140**
R1a	0.053	**0.318**	0.053	0	**0.106**
L	0.105	**0.168**	0.105	0.059	**0.109**
J2	0.105	0.121	**0.21**	0.059	**0.124**
C	0.053	0.03	0.053	0.059	**0.049**
K2	0.053	0.03	0	0	**0.021**
P*	0	**0.03**	0	0	**0.008**
F*	0	**0.015**	0	0	**0.004**
Haplogroup diversity (SE)	**0.842 (0.066)**	**0.834 (0.026)**	**0.836 (0.052)**	**0.809 (0.055)**	

Table 5. mtDNA diversity indices and demographic parameters in caste populations of western India.

	Caste populations			
	Maratha	**Dhangar**	**Desasth-brahmin**	**Chitpavan-brahmin**
	N = 19	**N = 19**	**N = 19**	**N = 20**
Diversity indices				
Polymorphic sites	70	61	57	53
Haplotype diversity (SE)	0.994 (0.019)	1.000 (0.017)	1.000 (0.017)	0.995 (0.017)
Nucleotide diversity (SE)	0.057 (0.030)	0.050 (0.027)	0.046 (0.025)	0.045 (0.023)
Mean Pairwise difference (π)	11.491 (5.451)	10.13 (4.84)	9.368 (4.502)	8.805 (4.239)
Demographic expansion parameters				
Fu's Fs	- 7.85 (p=0.004)	-11.24 (p=0)	-11.870 (p=0)	-10.424 (p<0)
Tajima' D	-1.8 (p=0.021)	-1.72 (p=0.027)	-1.74 (p=0.021)	-1.656 (p=0.032)
Raggednes index (r)	0.024	0.019	0.014	0.017
Tau (only HVS-1 region)	6.437	6.231	6.311	4.59
Expansion time based upon Tau value	54,183 years	52,450 years	53,123 years	38,636 years

Table 6. Analyses of Molecular Variance in caste and putative parental populations based upon uniparental and biparental markers

Groupings	% variation attributable to							
	Among groups				Among populations within groups			
	mtDNA[5, 6]	Y-SNP[5]	Y-STR	Autosomal STR	mtDNA[a,b]	Y-SNP[a]	Y-STR	Autosomal STR
4 Maharashtrian castes	NA	NA	NA	NA	1.71 (0.029)	5.29 (0.0009)	0.11 (0.240)	0.7 (0.0)
3 ranked castes: Brahmins, warrior and peasant	-1.76	-2.52	-0.1	0.47 (0.161)	3.18 (0.066)	7.16 (0.004)	0.19 (1.00)	0.29 (0)
2 Linguistic groups: Marathi and Konkani speakers	2.41 (0.25)	7.56 (0.27)	0.19 (0.268)	-0.33	0.60 (0.009)	-0.72	0.01 (0.344)	0.83 (0)
4 geographical groups (castes, CA, WA, EU) [1,2,3]	4.36 (0.001)	9.14 (0)	NA	NA	4.08 (p=0)	6.03 (p=0)	NA	NA
2 (castes & CA)	5.8 (0.034)	9.51 (0.033)	NA	NA	1.20 (0.021)	4.51 (0.001)	NA	NA
2 (castes & WA)	5.25 (0.03)	9.58 (0.03)	NA	NA	5.28 (0)	5.82 (0)	NA	NA
2 (castes & EU)	15.22 (0.03)	15.52 (0.03)	NA	NA	0.68 (0)	5.90 (0)	NA	NA

[1] CA: Central Asians, [2] WA: West Asians, [3] EU: Europeans;

[5] Kivisild et al. 2003; [6] Baig et al. 2004

Table 7. Admixture estimates based on bi- and uni-parental genetic markers in caste populations of western India

	% contribution [m (SE)] from putative parental populations*			
	Central Asians[1]	West Asians[2]	Europeans[3]	extant Maharashtrian castes
	N= 69	N=305	N= 133	N= 55 - 102
Chitpavan-brahmin:				
Y-SNP	1	0.125	**0.375**	-0.500
mt-DNA	-0.081	-0.064	0.084 (0.3)	1.062 (9.16)
autosomal STR	-0.0144	-0.084	**0.209 (0.06)**	0.89 (0.041)
Desasth-brahmin:				
Y-SNP	0.218	-0.062	0.28	0.56
mt-DNA	0.009 (9.7)	0.158 (4.52)	-0.158	0.972 (0)
autosomal STR	-0.003	0.010 (0.027)	0.212 (0.020)	0.78 (0.03)
Maratha:				
Y-SNP	0.273 (2.7)	**0.215**	-0.273	0.78 (6.5)
mt-DNA	**0.037 (9.8)**	-0.146	-0.291	**1.4 (7.6)**
autosomal STR	-0.063	0.42 (0.009)	0.03 (0.01)	0.994 (0.019)
Dhangar:				
Y-SNP	-0.625	0.065 (9.1)	0.048	**0.95**
mt-DNA	0.011	**0.238**	-0.216	0.97
autosomal STR	0.054 (0.005)	0.075 (0.001)	-0.47 (0.008)	1.34 (0.006)

* Table 1 for parental composition

[1] Turkish; [2] Ashkenazi-Jews, Iranian; [3] PT, GK, FR for autosomal STR loci

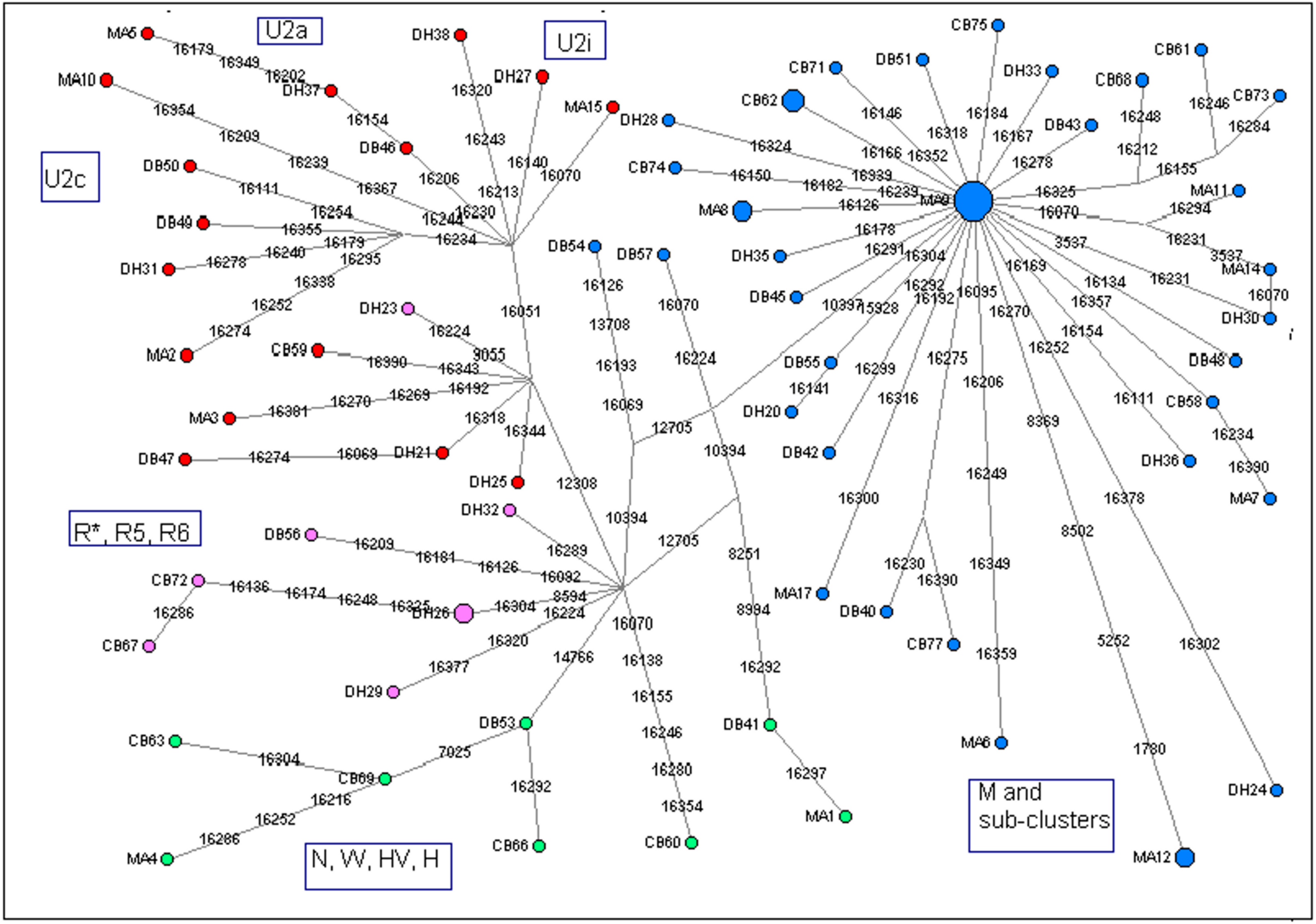

Figure 1

Figure 1

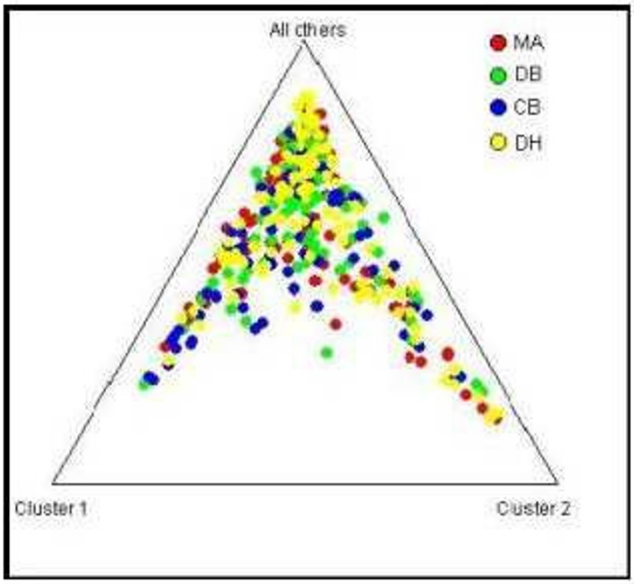

Figure 2

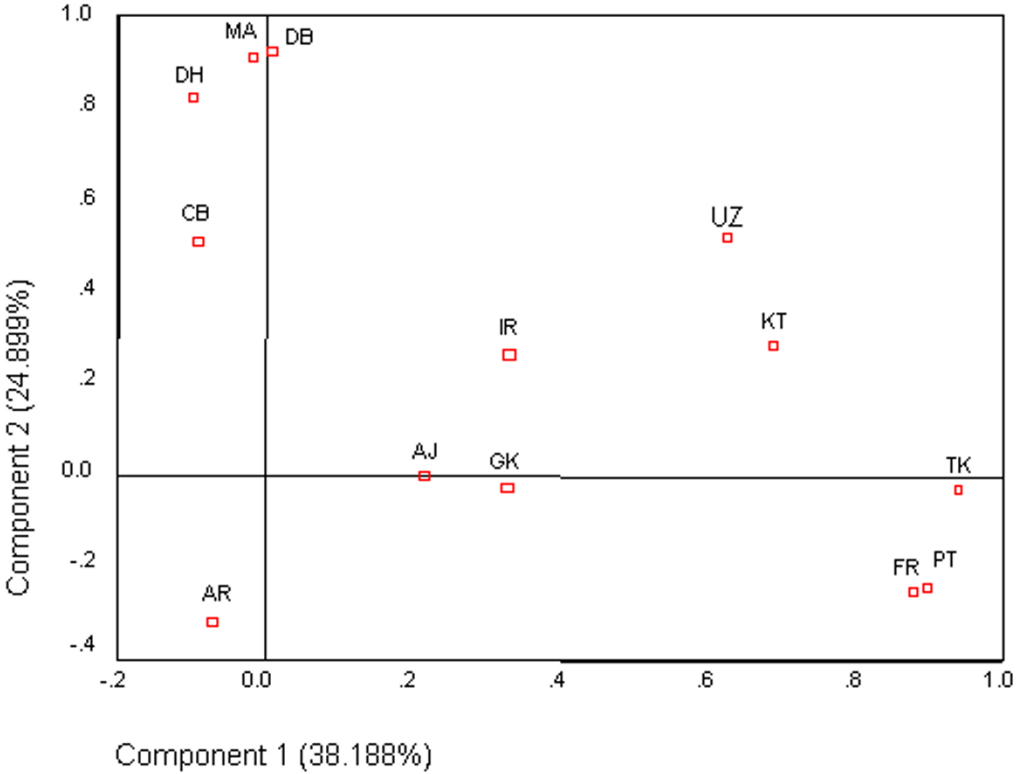

Figure 3a

Figure 3

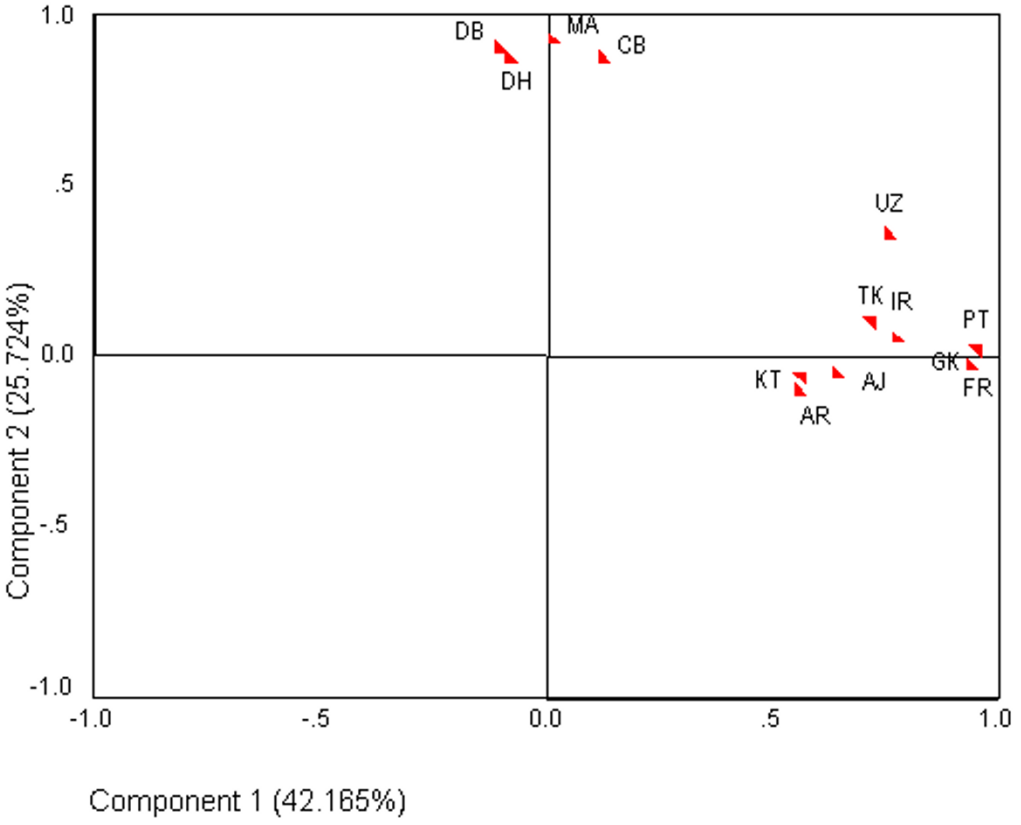

Figure 3b

Figure 4

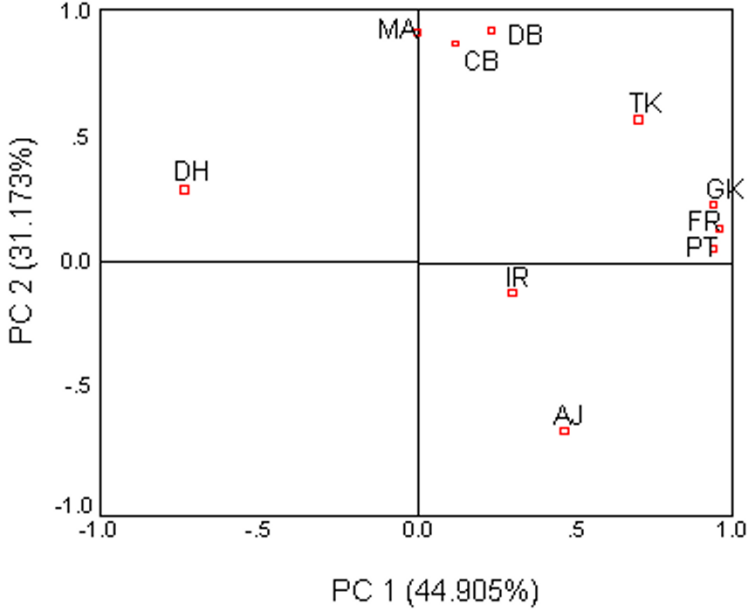

Figure 3c

Figure 5

Additional files provided with this submission:

Additional file 2 : Table 4 -additional file 2.xls : 169Kb
http://genomebiology.com/imedia/1995289715740764/sup2.XLS
Additional file 1 : Table 3 - additional file 1.xls : 42Kb
http://genomebiology.com/imedia/1328867408740764/sup1.XLS