Published in final edited form as:

Hum Genet. 2006 November ; 120(4): 543–551. doi:10.1007/s00439-006-0234-x.

A Shared Y-chromosomal Heritage between Muslims and Hindus in India

Ramana Gutala^{1,*}, Denise R. Carvalho-Silva^{2,*}, Li Jin³, Bryndis Yngvadottir², Vasanthi Avadhanula⁴, Khaja Nanne⁴, Lalji Singh⁵, Ranajit Chakraborty³, and Chris Tyler-Smith²

¹Department of Medicine, University of Texas Health Science Center, San Antonio, USA

²The Wellcome Trust Sanger Institute, Hinxton, UK

³Center for Genome information, University of Cincinnati, Cincinnati, USA

⁴Deccan College of Medical Sciences, Hyderabad, India

⁵Center for Cellular and Molecular Biology, Hyderabad, India

Abstract

Arab forces conquered the Indus Delta region in 711 A.D. and, although a Muslim state was established there, their influence was barely felt in the rest of South Asia at that time. By the end of the tenth century, Central Asian Muslims moved into India from the northwest and expanded throughout the subcontinent. Muslim communities are now the largest minority religion in India, comprising more than 138 million people in a predominantly Hindu population of over one billion. It is unclear whether the Muslim expansion in India was a purely cultural phenomenon or had a genetic impact on the local population. To address this question from a male perspective, we typed eight microsatellite loci and 16 binary markers from the Y chromosome in 246 Muslims from Andhra Pradesh, and compared them to published data on 4,204 males from China, Central Asia, other parts of India, Sri Lanka, Pakistan, Iran, the Middle East, Turkey, Egypt and Morocco. We find that the Muslim populations in general are genetically closer to their non-Muslim geographical neighbors than to other Muslims in India, and that there is a highly significant correlation between genetics and geography (but not religion). Our findings indicate that, despite the documented practice of marriage between Muslim men and Hindu women, Islamization in India did not involve large-scale replacement of Hindu Y chromosomes. The Muslim expansion in India was predominantly a cultural change and was not accompanied by significant gene flow, as seen in other places, such as China and Central Asia.

Keywords

Y-chromosomal polymorphism; India; Muslim; Hindu

Introduction

Islam is India's largest minority religion, with Muslims officially comprising ~13 % of the population, or 138 million people (Census 2001). The history of Islam in India began in the

Address for correspondence and reprints: Chris Tyler-Smith, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SA, UK., E-mail:cts@sanger.ac.uk, Tel: [+44] (0) 1223 495376, Fax: [+44] (0) 1223 494919.

*These two authors contributed equally to the article.

Electronic-Database Information

Census 2001, <http://www.censusindia.net/>

year 711 A.D., when it was introduced into Sind by the Arabs (Titus 2005). Soon after, however, Sind was abandoned and for the next two and a half centuries there was little Muslim presence in India. Then, in 1001, the Turks entered India from Afghanistan and started spreading Islam from west to east (Titus 2005). By the beginning of the fourteenth century the Deccan in south India had been invaded, and soon after that the Muslim empire and influence attained its greatest extent and importance in the history of India, remaining dominant up to 1707 A.D. (Titus 2005).

The Muslim conquest of India was undertaken with the purpose of establishing a Muslim government over the people and implementing the Muslim faith. This was accomplished by foreign conquerors, traders, religious devotees and preachers using a wide range of methods, including war, enslavement and conversion (voluntary or compulsory), and through marriage between Muslims and Hindus (Lal 1993; Titus 2005). Such mixed marriages appear to have been part of the policy of absorption and domination by which it was hoped Hinduism would be overthrown (Titus 2005). For that reason, the practice became well established and the resulting progeny contributed extensively to the increase in the Muslim populations in India (Lal 1993; Titus 2005).

The biological contribution to India accompanying these historical events has not been thoroughly investigated and the extensive studies of Indian genetic pre-history (reviewed by McElreavey and Quintana-Murci 2005) have focused on the origin of caste and tribal populations, the birthplace of the Dravidian languages, and the contribution of genes from the Indo-European speakers during their movement out of Central Asia (e.g. Sahoo et al. 2006; Sengupta et al. 2006). The few studies examining the origins of Muslims in India have provided conflicting results. Classical marker studies, for example, have shown that Muslims and Hindus in north and northwestern India are different from each other (Aarzo and Afzal 2005; Balgir 2003; Balgir and Sharma 1988), whereas a study of the Y chromosome revealed close affinity between Muslims and Indo-European upper-caste groups (Basu et al. 2003). Since the expansion of Islam in other places, such as China and Central Asian countries, involved the movement of people and Y chromosomes (Wang et al. 2003; Zerjal et al. 2002) and left a detectable genetic signature in the current populations, a similar genetic impact from the Middle East on the Hindu gene pool seems plausible, but needs further investigation.

We therefore set out to clarify this aspect of the history of India by studying 24 Y-chromosomal markers in 246 Muslims from south India, and comparing our results to published data on 4,204 Muslim and non-Muslim males from several other countries. By investigating a large set of Indian Muslims and performing a comprehensive analysis of the data, we show that in India the spread of Islam did not have a detectable genetic impact on the local populations and thus differed from its expansion in neighboring countries. In India, the spread of Islam was predominantly a cultural event.

Subjects and Methods

DNA samples and Y-chromosomal polymorphisms

The sample consisted of 246 unrelated males from five different populations from Andhra Pradesh, South India: Yamani, Pathans and Bohra Muslim groups, and two other Sunni and Shia groups here referred to as “Sunni” and “Shia”, respectively. Blood samples were collected with informed consent and DNA was extracted following standard procedures. Eight microsatellite loci (*DYS19*, *DYS388*, *DYS389I*, *DYS389II*, *DYS390*, *DYS391*, *DYS392* and *DYS393*) and 16 biallelic markers (YAP, M9, M89, M52, M45, M173, M172, M17, M11, M15, M40, M70, M147, M95, M103 and M88) were typed as previously described (Ramana et al. 2001) and used to assign haplotypes and Y haplogroups,

respectively. In addition, relevant Y-chromosomal data from literature sources were collated and analyzed. In compiling these data, we were unable to reconcile all *DYS389* repeat counts from different sources satisfactorily, and so excluded this locus from our analyses. Data from 4,204 males (non-Muslims and Muslims) from other parts of India, China, Central Asia, Sri Lanka, Pakistan, Iran, the Middle East, Turkey, Egypt, and Morocco were included (Fig. 1, Table 1).

Data analyses

Both haplotype and haplogroup frequencies were determined, and combined with their molecular information to compute genetic distances between all the populations depicted in Fig. 1. Pairwise distances based on microsatellite markers (R_{st}) and on biallelic marker (Φ_{st}) were obtained with Arlequin 2.0 (Schneider et al. 2000). Distance matrices separating each pair of populations were then used to perform multidimensional scaling (MDS) analysis with the SPSS 13.0 software package. Negative genetic distances were assigned a value of zero; when we alternatively increased all distances to eliminate the negative values, or used additional software tools (Statistica 6), the results were very similar (not shown). For the Indian samples only, we combined the populations into classes and computed average R_{st} values 1) among Muslims, 2) among non-Muslims, and 3) between Muslims and non-Muslims using a jackknife approach within each group. Mantel tests to assess the significance of correlations between genetics and religion, or geography were carried out in populations from India by use of Arlequin. Analysis of molecular variance (AMOVA) was also performed with Arlequin using microsatellite and biallelic data in Indian populations, which were either grouped according to religion (Muslims and non-Muslims) or geographical regions, or not grouped at all. The possibility of gene flow among the different Muslim isolates, and among Muslims and Hindus in south India, was investigated by estimating the proportion of lineage sharing and the rho genetic distance (Helgason et al. 2000).

Results and Discussion

We typed eight Y-chromosomal microsatellites and 16 binary markers in 246 Muslim men from Andhra Pradesh (south India), and defined 124 different haplotypes, or five haplogroups and four paragroups, respectively (Supplementary Table 1). We then compared our data (excluding *DYS389*) to published data from 4,204 males (Muslims and non-Muslims) from other parts of India, China, Central Asia, Sri Lanka, Pakistan, Iran, the Middle East, Turkey, Egypt, and Morocco (Table 1, Fig.1).

For this worldwide comparison, R_{st} and Φ_{st} genetic distances were calculated between all the populations and their pairwise values were used to perform an MDS analysis. The resulting plots (Fig. 2) showed considerable structure. Although a continuum of variation is seen, rather than discrete groups, populations from a particular region or country tend to cluster together; this is in agreement with the expectation that human genetic structure is predominantly geographical and clinal. Thus, for example, most Chinese populations are seen in the left-hand part of each plot rather than dispersed throughout the plot. Interestingly, however, the three Chinese Muslim populations do not lie in this cluster, but are located more towards the centre of each plot, close to populations with geographical origins lying further west. It has previously been reported that the conversion to Islam in China involved the movement of people, and, in particular, the influx of genes from the Middle East into China (Wang et al. 2003) and Figure 2 thus confirms that our analysis readily detects such events. The Indian populations show considerable diversity, and northern and southern populations barely overlap (with the exception of Dravidians and Chenchu) and tend to lie in two distinct clusters (Fig. 2). Most importantly, the Muslim and non-Muslim populations are intermingled in these clusters: the Y-chromosomal heritage in

India is influenced more by geographical location than by the religious practices. It appears that the Muslim genetic contribution in India was less important than in other places such as China.

In order to assess the significance of this observation, we next combined the populations from India into two classes, Muslims and non-Muslims, and calculated pairwise genetic distances 1) among Muslims, 2) among non-Muslims, and 3) between Muslims and non-Muslims, and compared their average values after following a jackknife approach within each group. The comparison between Muslims and non-Muslims in India showed the lowest distance (Fig. 3). We then restricted the comparisons to Muslim and non-Muslim populations who live in neighboring regions of south India (again under a jackknife approach). We found that this comparison resulted in the lowest average value of genetic distances (Fig. 3), which suggests that the close geographic proximity of Muslims and non-Muslims in south India might have facilitated gene flow between those two groups. Our hypothesis of geography playing a more important role than religion in structuring Y-chromosomal diversity in India was then assessed by means of a Mantel test. This test asks whether there is a correlation between geographic distances (or religion) and genetic distances. Genetic distances were based on R_{st} or Φ_{st} , geographic distances were calculated using the approximate latitude and longitude of the sample sites, and religious distances were defined as 0 or 1 according to whether or not the populations belonged to the same religious group. Figure 4 shows that when 19 Indian populations are considered (Table 1), there is a correlation between genetic distances and geographic distances ($r=0.43$, $p<0.001$ for microsatellites; $r^2=0.24$, $p<0.01$, for biallelic markers) but not between genetics and religion ($r=0.10$, $p>0.05$ for microsatellites; $r^2=0.08$, $p>0.05$, for biallelic markers). The correlation is even stronger when the test is performed in populations from north India versus south India only, ($r=0.63$, $p<0.001$, for microsatellites; $r^2=0.50$, $p<0.01$, for biallelic markers), but not with religion ($r=0.03$, $p>0.05$, for microsatellites; $r^2=0.03$, $p>0.05$, for biallelic markers). This positive correlation between the Y diversity and geography still remains when the same test is performed among populations from south India only, despite the shorter geographic distances between them ($r=0.16$, $p<0.05$, for microsatellite data only) (Fig. 4). Thus these results indicate that in India, the processes that cause a positive correlation between the pattern of Y variation and geography are not disrupted by religious affiliation. It is also worth noting that stronger support is obtained with Y-chromosomal microsatellites than with biallelic markers, which may reflect the less biased measure of diversity provided by microsatellites.

We then performed an analysis of molecular variance (AMOVA) using both microsatellite and biallelic markers in 19 populations from north India, south India and from the remaining regions of India (Table 2A); and in 14 populations from north India and south India only (Table 2B). As expected, the highest fraction of variation was within populations when no grouping was defined (Table 2). We then pooled the populations into two groups according to religion (Muslim or non-Muslim) or geography. With the first grouping, the amount of variation among populations from the same group was always higher than the among-group variation (Table 2). However, when we grouped the populations according to the geographic regions in India, the fraction of variation among groups was significantly higher than the among-population within-group variation, and ranged from 8.2 to 12.8% depending on the regions and markers considered (Table 2). These results confirm the large differences between populations that live in south India and those that live in north India, rather than between Muslims and non-Muslims. We also performed AMOVA in nine populations from south India only and pooled them in two groups according to their religion, i.e. Muslims and non-Muslims. The among-population variation in south India only was 5.4%, lower than the values of 9.2% obtained when considering Muslims and non-Muslims from larger geographic regions of India (Table 2). Overall, the AMOVA analyses emphasize the

importance of geography in shaping the Y diversity in India and give further support to our hypothesis of no major contribution of Muslim Y chromosomes into the Hindu paternal gene pool during the Islamization of India.

Finally, we assessed the evidence of gene flow among the different south Indian Muslim isolates and among Muslims and Hindus in south India. We calculated both the proportion of shared haplotypes in the two sets of two populations, and the rho distance (the average number of mutations between a haplotype in one population and its closest counterpart in the second population) (Helgason et al. 2000). These measures are more sensitive to low levels of gene flow, but were not significantly different between Muslims and Hindus (Table 3), confirming the lack of genetic differentiation according to religious affinity in India.

Although marriage between Muslim men and Hindu women was important for the spread of Islam in India, it has not been sufficient to replace the Hindu Y-chromosomal heritage built up in prehistoric times. This is in contrast with observations in Muslim groups from other places such as China and Central Asia, where there has been more marked movement of Muslim Y chromosomes into the area. Our conclusion does assume that the Muslim population entering India would have been genetically distinct from the indigenous populations, which seems likely in view of their distinct geographical origin. Moreover, our results are in accordance with previous work on the sharing of Y-chromosomes among different religious communities that live side by side, namely Jewish groups and their non-Jewish neighbors in the Near East (Hammer et al. 2000; Nebel et al. 2000; Thomas et al. 2002).

At least at the Y-chromosomal level, the origin of Muslim isolates in south India is predominantly from local populations rather than from other Muslims of other parts of India, or outside the country. Some Indian Muslim families can trace their ancestry back to sources outside India >1,000 years ago, and our findings do not conflict with this fact, but do show that the largest minority religious group in India arose in the main from a cultural change among Hindus who started to follow and spread the precepts of Islam. The Y-chromosomal variation among Indian populations reflects geographical and prehistorical factors rather than the practices of Hinduism or Islam.

Acknowledgments

We specially thank all the donors for making this work possible; George van Driem for encouragement; Toomas Kivisild, Sarabjit Mastana, Partha P. Majumder, Peter Underhill and Rene Herrera for helpful information; Joan Green and Andrew King for facilitating the access to historical books; S. Qasim Mehdi, Tatiana Zerjal and Oscar Lao for comments and discussions; and three referees for suggesting improvements to the manuscript. DRC-S was supported by funds from the Arts and Humanities Research Board and the EC Sixth Framework Programme under Contract no. ERAS-CT-2003-980409. CT-S was supported by The Wellcome Trust.

References

- Aarzo SS, Afzal M. Gene diversity in some Muslim populations of North India. *Hum Biol.* 2005; 77:343–53. [PubMed: 16392636]
- Balgir RS. Morphological and regional variations in body dimensions of the Gujjars of different localities in north-western India. *Anthropol Anz.* 2003; 61:275–85. [PubMed: 14524001]
- Balgir RS, Sharma JC. Genetic markers in the Hindu and Muslim Gujjars of Northwestern India. *Am J Phys Anthropol.* 1988; 75:391–403. [PubMed: 3163234]
- Basu A, Mukherjee N, Roy S, Sengupta S, Banerjee S, Chakraborty M, Dey B, Roy M, Roy B, Bhattacharyya NP, Roychoudhury S, Majumder PP. Ethnic India: a genomic view, with special reference to peopling and structure. *Genome Res.* 2003; 13:2277–90. [PubMed: 14525929]

- Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA. Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet.* 2004; 114:127–48. [PubMed: 14586639]
- Hammer MF, Redd AJ, Wood ET, Bonner MR, Jarjanazi H, Karafet T, Santachiara-Benerecetti S, Oppenheim A, Jobling MA, Jenkins T, Ostrer H, Bonn-Tamir B. Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. *Proc Natl Acad Sci U S A.* 2000; 97:6769–74. [PubMed: 10801975]
- Helgason A, Sigurardttir S, Nicholson J, Sykes B, Hill EW, Bradley DG, Bosnes V, Gulcher JR, Ward R, Stefnsson K. Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. *Am J Hum Genet.* 2000; 67:697–717. [PubMed: 10931763]
- 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, Vilems R. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet.* 2003; 72:313–32. [PubMed: 12536373]
- Lal, KS. *Indian Muslims: Who are they.* Voice of India; New Delhi, India: 1993.
- Luis JR, Rowold DJ, Regueiro M, Ciro B, Cinnioglu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ. The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am J Hum Genet.* 2004; 74:532–44. [PubMed: 14973781]
- McElreavey K, Quintana-Murci L. A population genetics perspective of the Indus Valley through uniparentally-inherited markers. *Ann Hum Biol.* 2005; 32:154–62. [PubMed: 16096211]
- 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; 69:1095–112. [PubMed: 11573163]
- Nebel A, Filon D, Weiss DA, Weale M, Faerman M, Oppenheim A, Thomas MG. High-resolution Y chromosome haplotypes of Israeli and Palestinian Arabs reveal geographic substructure and substantial overlap with haplotypes of Jews. *Hum Genet.* 2000; 107:630–41. [PubMed: 11153918]
- Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi SQ. Y-chromosomal DNA variation in Pakistan. *Am J Hum Genet.* 2002; 70:1107–24. [PubMed: 11898125]
- 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. Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am J Hum Genet.* 2001; 68:537–42. [PubMed: 11133362]
- Ramana GV, Su B, Jin L, Singh L, Wang N, Underhill P, Chakraborty R. Y-chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe, and the migrant Siddi populations of Andhra Pradesh, South India. *Eur J Hum Genet.* 2001; 9:695–700. [PubMed: 11571559]
- Sahoo S, Singh A, Himabindu G, Banerjee J, Sitalaximi T, Gaikwad S, Trivedi R, Endicott P, Kivisild T, Metspalu M, Vilems R, Kashyap VK. A prehistory of Indian Y chromosomes: Evaluating demic diffusion scenarios. *Proc Natl Acad Sci U S A.* 2006; 103:843–8. [PubMed: 16415161]
- Schneider, S.; Roessli, D.; Excoffier, L. *Arelquin: a software for population genetics data analysis.* 2.000 edn.. Genetics and Biometry Lab., Dept. of Anthropology, University of Geneva; 2000.
- Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CE, Lin AA, Mitra M, Sil SK, Ramesh A, Usha Rani MV, Thakur CM, Cavalli-Sforza LL, Majumder PP, Underhill PA. Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am J Hum Genet.* 2006; 78:202–21. [PubMed: 16400607]
- Thomas MG, Weale ME, Jones AL, Richards M, Smith A, Redhead N, Torroni A, Scozzari R, Gratrix F, Tarekegn A, Wilson JF, Capelli C, Bradman N, Goldstein DB. Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. *Am J Hum Genet.* 2002; 70:1411–20. [PubMed: 11992249]
- Titus, MT. *Islam in India and Pakistan.* Munshiram Manoharlal Publishers; New Delhi, India: 2005.
- Wang W, Wise C, Baric T, Black ML, Bittles AH. The origins and genetic structure of three co-resident Chinese Muslim populations: the Salar, Bo'an and Dongxiang. *Hum Genet.* 2003; 113:244–52. [PubMed: 12759817]

- Xue Y, Zerjal T, Bao W, Zhu S, Shu Q, Xu J, Du R, Fu S, Li P, Hurles ME, Yang H, Tyler-Smith C. Male demography in East Asia: a north-south contrast in human population expansion times. *Genetics*. 2006; 172:2431–9. [PubMed: 16489223]
- Zerjal T, Wells RS, 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; 71:466–82. [PubMed: 12145751]

Figure 1.

Geographical locations of the populations analyzed. Symbol shapes indicate religion (squares for Muslims and circles for non-Muslims). The colours represent locations: Morocco (light blue), Egypt (orange), Israel (black), Iran, Iraq and Oman (yellow), Turkey, Azerbaijan, Georgia and Armenia (red), Kyrgyzstan, Kazakhstan, Uzbekistan and Turkmenistan (white), China, Korea, Mongolia and Japan (brown), Pakistan (dark green), North India (light green), South India (dark blue), Sri Lanka (pink).

Figure 2.

Multidimensional scaling presentation of population pairwise values of R_{st} and ϕ_{st} based on Y-microsatellite haplotypes (A) and Y-biallelic markers (B). Symbol shapes indicate religion (squares for Muslim and circles for non-Muslims). RSQ value (stress value) = 0.89 (0.16) and 0.86 (0.18) for microsatellites and biallelic markers, respectively. Population colour codes are as in Fig. 1. Two south Indian populations (Dravidian and Chenchu) clustered with populations from north India and were not considered in the remaining analyses.

Figure 3.

The averaged genetic distances (R_{st}) after following a jackknife approach between groups of populations based on 6 Y-microsatellites among Muslims (pattern filled), among non-Muslims (white), between Muslims and non-Muslims (black) in India and between Muslims and non-Muslims (grey) in south India.

Figure 4. Correlation coefficient between genetics and geography (white) and genetics and religion (black) in populations from south, north and the remaining regions of India (1), in populations from south and north India (2), and in populations from south India only (3), based on Y-microsatellite haplotypes (A) and biallelic marker haplogroups (B). ** $p < 0.001$, * $p < 0.05$.

Table 1

Muslim and non-Muslim populations studied

Population	Number of samples typed		Country/region	Reference
	microsatellites	biallelic makers		
Bohra ^a	48	45	India/South	Present work
Pathans	50	50	India/South	Present work
Shia	49	47	India/South	Present work
Sunni	50	50	India/South	Present work
Yamani	49	49	India/South	Present work
Kamma	39	40	India/South	(Ramana et al. 2001)
Peruru	43	44	India/South	(Ramana et al. 2001)
Poroja	20	20	India/South	(Ramana et al. 2001)
Valmiki	24	24	India/South	(Ramana et al. 2001)
Vizag Brahmin	41	41	India/South	(Ramana et al. 2001)
Dravidian (caste)	66	103	India/South	(Basu et al. 2003)
Chenchu	41	42	India/South	(Kivisild et al. 2003)
Lambadi	40	41	India/South	(Kivisild et al. 2003)
Brahmin	15	17	India/North	(Basu et al. 2003)
Chamar	18	18	India/North	(Basu et al. 2003)
Muslim	16	19	India/North	(Basu et al. 2003)
Rajput	34	35	India/North	(Basu et al. 2003)
Brahmin	43	44	India/West	(Kivisild et al. 2003)
Gujarat	29	29	India/West	(Kivisild et al. 2003)
Punjab	60	66	India/Northeast	(Kivisild et al. 2003)
Indo-European (tribe)	16	19	India/Central	(Basu et al. 2003)
West Bengal	31	31	India/East	(Kivisild et al. 2003)
Singal	30	42	SriLanka	(Kivisild et al. 2003)
Moor	46	47	SriLanka	(Kivisild et al. 2003)
Baluch	59	59	Pakistan/Baluchistan	(Qamar et al. 2002)
Brahui	110	110	Pakistan/Baluchistan	(Qamar et al. 2002)
Burusho	94	94	Pakistan/Karakorum Mountains	(Qamar et al. 2002)
Hazara	23	23	Pakistan/Southern Baluchistan	(Qamar et al. 2002)
Kalash	44	44	NWFP	(Qamar et al. 2002)
Makrani Baluch	25	28	Makran coast	(Qamar et al. 2002)
Makrani Negroid	33	33	Makran coast	(Qamar et al. 2002)
Parsi	90	90	Karachi	(Qamar et al. 2002)
Pathan	93	93	NWFP and Baluchistan	(Qamar et al. 2002)
Sindhi	122	122	Sindhi	(Qamar et al. 2002)
Daur	39	39	China/Northeast	(Xue et al. 2006)
Ewenki	26	27	China/Northeast	(Xue et al. 2006)

Population	Number of samples typed		Country/region	Reference
	microsatellites	biallelic makers		
Hezhe	44	45	China/Northeast	(Xue et al. 2006)
Hui	35	35	China/Central	(Xue et al. 2006)
Manchu	35	35	China/East	(Xue et al. 2006)
Inner Mongolian	45	45	China/Northeast	(Xue et al. 2006)
Oroqen	31	31	China/Northeast	(Xue et al. 2006)
Uygur_1 (Urumqi)	31	31	China/Northwest	(Xue et al. 2006)
Uygurs_2 (Yili)	39	39	China/Northwest	(Xue et al. 2006)
Xibe	41	41	China/Northwest	(Xue et al. 2006)
Han (Heilongjiang)	35	35	China/Northeast	(Xue et al. 2006)
Han (Xinjiang)	32	32	China/Northwest	(Xue et al. 2006)
Chinese Korean	25	25	China/Northeast	(Xue et al. 2006)
Buyi	35	35	China/South	(Xue et al. 2006)
Hani	34	34	China/South	(Xue et al. 2006)
Li ^a	34	34	China/South	(Xue et al. 2006)
Qiang	33	33	China/Central	(Xue et al. 2006)
She	34	34	China/Southeast	(Xue et al. 2006)
Tibetans	35	35	China/Southwest	(Xue et al. 2006)
Yao (Bama)	35	35	China/South	(Xue et al. 2006)
Yao (Liannan)	35	35	China/South	(Xue et al. 2006)
Han (Sichuan)	34	35	China/South	(Xue et al. 2006)
Han (Gansu)	30	30	China/Central	(Xue et al. 2006)
Han (Guangdong)	35	35	China/Southeast	(Xue et al. 2006)
Japanese	47	47	Japan	(Xue et al. 2006)
Korean	43	42	Korean	(Xue et al. 2006)
Outer Mongolian	58	65	Mongolia	(Xue et al. 2006)
Kyrgyz ^a	37	41	Kyrgyzstan	(Zerjal et al. 2002)
Dungans	22	22	Kyrgyzstan	(Zerjal et al. 2002)
Uyghurs	33	33	Kazakhstan	(Zerjal et al. 2002)
Kazaks	14	38	Kazakhstan	(Zerjal et al. 2002)
Uzbeks	29	28	Uzbekistan	(Zerjal et al. 2002)
Tajiks ^a	22	22	Tajikistan	(Zerjal et al. 2002)
Turkmen	21	21	Turkmenistan	(Zerjal et al. 2002)
Kurds	20	20	Turkmenistan	(Zerjal et al. 2002)
Georgians	26	26	Georgia	(Zerjal et al. 2002)
Ossetians	15	15	Georgia	(Zerjal et al. 2002)
Lezgi	12	12	Azerbaijan	(Zerjal et al. 2002)
Svans ^a	25	25	Georgia	(Zerjal et al. 2002)
Azeri	19	19	Azerbaijan	(Zerjal et al. 2002)

Population	Number of samples typed		Country/region	Reference
	microsatellites	biallelic makers		
Armenians	21	21	Armenia	(Zerjal et al. 2002)
Kurds-Iran	18	18	Iran/Western	(Quintana-Murci et al. 2001)
Mazandarians	18	18	Iran/Northern	(Quintana-Murci et al. 2001)
Gilaks	38	38	Iran/Northern	(Quintana-Murci et al. 2001)
Lurs	16	16	Iran /Southwestern	(Quintana-Murci et al. 2001)
Persians	51	51	Iran/Central and southern central	(Quintana-Murci et al. 2001)
Turks	40	40	Azerbaijan/Eastern and western	(Quintana-Murci et al. 2001)
Muslim Kurds	95	95	Iraq/Northern	(Nebel et al. 2001)
Ashkenazi	79	79	Israel	(Nebel et al. 2001)
Shephardic	78	78	Israel	(Nebel et al. 2001)
Kurdish Jew	99	99	Israel	(Nebel et al. 2001)
Palestinian Arabs	143	143	Israel/Palestine	(Nebel et al. 2001)
Beduins ^a	32	32	Israel/South	(Nebel et al. 2001)
Turkish 1	50	50	Turkey/Northeast	(Cinnioglu et al. 2004)
Turkish 2	27	27	Turkey/North	(Cinnioglu et al. 2004)
Turkish 3	76	77	Turkey/Northeast	(Cinnioglu et al. 2004)
Turkish 4	67	67	Turkey/East	(Cinnioglu et al. 2004)
Turkish 5	36	36	Turkey/Southeast	(Cinnioglu et al. 2004)
Turkish 6	31	31	Turkey/South	(Cinnioglu et al. 2004)
Turkish 7	86	86	Turkey/Central	(Cinnioglu et al. 2004)
Turkish 8	29	29	Turkey/West	(Cinnioglu et al. 2004)
Turkish 9	75	76	Turkey/Istambul	(Cinnioglu et al. 2004)
Tanta	120	147 ^b	Egypt	(Luis et al. 2004)
Cairo	28		Egypt	(Luis et al. 2004)
Oman	119	121	Oman	(Luis et al. 2004)
Arabs	60	no SNP data	Morocco	(Quintana-Murci et al. 2001)
Berbers ^a	49	no SNP data	Morocco	(Quintana-Murci et al. 2001)
Total	4447	4298		

^aThese populations were excluded from the main analyses because of their low genetic diversity and extreme pairwise values of genetic distances. Their geographic location is nevertheless shown in Fig.1.

^bThis number represents the samples collected in Egypt (both Cairo and Tanta locations)

Table 2

AMOVA results

	microsatellites				biallelic markers			
	Variance	Φ_{ST}	Φ_{SC}	Φ_{CT}	Variance	Φ_{ST}	Φ_{SC}	Φ_{CT}
A) North India, south India, remaining regions								
No grouping:	(%)				(%)			
Among populations	9.6				9.02			
Within populations	90.4	0.096			90.98	0.09		
Religion:								
Among groups	0.75			0.008 ^a	0.87			0.01 ^a
Among populations within groups	9.21		0.09		8.58		0.09	
Within populations	90.03	0.099			90.55	0.09		
Geography:								
Among groups ^a	8.17			0.082	8.7			0.087
Among populations within groups	5.07		0.06		3.48		0.04	
Within populations	86.76	0.132			87.81	0.12		
B) North India and south India								
No grouping:	(%)				(%)			
Among populations	8.27				6.3			
Within populations	91.73	0.083			93.7	0.06		
Religion:								
Among groups	-0.47			-0.01 ^a	-0.57			-0.01 ^a
Among populations within groups	9.25		0.09		8.77		0.09	
Within populations	91.22	0.088			91.8	0.08		
Geography:								
Among groups	9.48			0.095	12.83			0.128
Among populations within groups	5.21		0.06		3.59		0.04	
Within populations	85.31	0.147			83.58	0.16		

NOTE: All *p* values are <0.01, except when stated^a *p* > 0.2

Table 3

Comparison of Y-STR Lineages in South India

Comparison	Lineage sharing ^a	ρ Distance ^b
Non-Muslims versus Muslims	0.06 $\pm$ 0.03	1.53 $\pm$ 0.31
Muslims versus Muslims	0.08 $\pm$ 0.02	1.69 $\pm$ 0.86

^aProportion of all lineages in a pair of populations that are shared between the populations, averaged over all population pairs (mean and SD)

^b ρ distance between pairs of haplotypes, averaged over all pairs in the populations (mean and SD)