 Int J Hum Genet, 8(1-2): 21-29 (2008)

The Grandest Genetic Experiment Ever Performed on Man? – A Y-Chromosomal Perspective on Genetic Variation in India

Denise R. Carvalho-Silva and Chris Tyler-Smith

The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambs. CB10 1SA, UK

KEYWORDS Y chromosome; genetic variation; Indian caste system; endogamy; population substructure

ABSTRACT We have analysed Y-chromosomal data from Indian caste, Indian tribal and East Asian populations in order to investigate the impact of the caste system on male genetic variation. We find that variation *within* populations is lower in India than in East Asia, while variation *between* populations is overall higher. This observation can be explained by greater subdivision within the Indian population, leading to more genetic drift. However, the effect is most marked in the tribal populations, and the level of variation between caste populations is similar to the level between Chinese populations. The caste system has therefore had a detectable impact on Y-chromosomal variation, but this has been less strong than the influence of the tribal system, perhaps because of larger population sizes in the castes, more gene flow or a shorter period of time.

INTRODUCTION

"The caste system in India was the grandest genetic experiment ever performed on man" wrote Theodosius Dobzhansky in his book *Genetic Diversity and Human Equality* (1973, page 31). The wording – 'man' instead of 'human' – now seems outdated, but perhaps remains applicable to this review since it will be restricted to the male-specific variation carried by the Y chromosome. What were the genetic consequences of the caste system for Y-chromosomal variation? Of course, every experiment requires a control. A 'control' for this 'experiment' would need to be a population of similar size that does not have a caste system. In practice, caste populations can be compared with the somewhat less numerous non-caste populations in India or the slightly more numerous populations in the adjacent region of East Asia – China and its neighbours.

In this review, we therefore begin by considering the relevant properties of the caste system and Y-chromosomal genetics in order to identify effects that we might look for. We then consider the datasets from India, China and other nearby countries that are available in the literature. Finally, we present new analyses of these data and discuss the insights they provide into the comparative male genetics of India and East Asia, and the limitations of the conclusions that can be drawn.

Corresponding author: Chris Tyler-Smith,
The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambs. CB10 1SA, UK.
Telephone: [+44] (0) 1223 495376,
Fax: [+44] (0) 1223 494919
E-mail: cts@sanger.ac.uk

EXPECTATIONS: THE RELATIONSHIP BETWEEN THE CASTE SYSTEM AND Y-CHROMOSOMAL VARIATION

The caste system divides society into endogamous groups. Key issues for this review are:

- What size are the groups?
- How strict is the endogamy?
- How long has the system been in existence?

None of these questions is easy to answer in a precise way. The 2001 census provided a figure of ~1,028,700,000 for the population of India (Census of India 2001) while the People of India project has identified 4,635 communities (Singh 1993), suggesting an average size of around 220,000 for each of these communities. These communities can even be thought of as occupying distinct ecological niches (Gadgil and Malhotra 1983). However, the variation in size between different communities is enormous, and the communities defined in this project are not necessarily equivalent to the endogamous groups that the geneticist would be interested in. Nevertheless, these figures show that the Indian population is socially highly substructured.

Gene flow between castes is rare, and when it does occur consists principally of hypergamy, where a woman marries a man of higher caste

and is absorbed into the new caste (Misra 2001). This does not result in any movement of Y chromosomes between castes. The equivalent practice for men, in which a man would marry a woman of higher caste and be absorbed into the higher caste (and a Y chromosome would thus move between castes), appears not to have been documented (Bhattacharyya et al. 1999). On the basis of social rules and historical records, therefore, Y chromosomes would be expected to remain strictly within their castes. Genetic data can provide independent insights into the level of undocumented hyperandry and interpretations have varied from low levels to the possibility of quite high levels (Reddy et al. 2005; Wooding et al. 2004; Zerjal et al. 2007).

The origins of the caste system are associated with the entry of Indo-Aryan speakers ~3,500 years ago (Thapar 1990; Wolpert 1997). Fortunately, a significant source of information about their society is available in the form of the *Rig-Veda*, a collection of over 1,000 hymns dating perhaps from as early as 1,500 BC. Indo-Aryan tribal society was organised into priests, warriors and commoners who formed the basis of the *Brahmin*, *Kshatriya* and *Vaishya* castes, with a fourth, *sudras*, added in India and further developments occurred later. One line from the *Rig-Veda* illustrates the fluidity of the early caste boundaries: "*I am a poet, my father is a physician and my mother is a grinder of corn*", and there is debate about how rigid the system has really been over long periods of history (Thapar 1990). The caste system was abolished by the Government of India in 1949. Nevertheless, 3,500 years would represent ~117 generations at 30 years per generation and provide a timescale over which significant genetic changes could accumulate.

The properties of the Y chromosome that make it particularly suitable for such analyses have been reviewed elsewhere (e.g. Jobling and Tyler-Smith 2003) and need only a brief mention here. In addition to its male-specific inheritance, the lack of recombination over most of the length of the chromosome results in long stable haplotypes, which change only by accumulating mutations. The abundant Single Nucleotide Polymorphism (SNP) and Short Tandem Repeat (STR) markers now available allow these to be characterised in detail. As a result, haplotypes can be both clustered into haplogroups that usually reflect shared ancestry thousands or tens of thousands of years ago, and resolved into family-specific (if not individual-specific) haplotypes. In addition, the large variance in the number of children fathered by different men results in strong genetic drift, leading to large differences between populations.

If the influence of natural selection on Y-chromosomal haplotypes can be ignored (Jobling and Tyler-Smith 2000), the pattern of variation found within a set of populations that are largely isolated from one another will be dominated by loss of variation due to random genetic drift, counterbalanced to some extent by increases due to gene flow and mutation. The amount of genetic drift is measured by the long-term effective population size, which depends on the census number of males in the population, the proportion who father children, the variance in number of children, the generation time and the extent to which these factors are correlated between generations. While some of these factors can readily be measured or modelled, others, such as the correlations between generations, are poorly understood. We therefore take an empirical approach and consider next some examples of isolated populations outside India as guides to the amount of Y-chromosomal drift that can be found in different circumstances.

Tristan da Cunha lies in the South Atlantic Ocean and has been described as 'the remotest island in the world'. Its population was established in 1816 by seven females and eight males, and currently numbers 269 with seven surviving surnames (Wikipedia: Tristan da Cunha 2007). A genetic survey published in 2003 identified eight main Y-chromosomal lineages, and a one-STR-step variant of one (Soodyall et al. 2003). Seven of these corresponded to seven of the eight founding males; the eighth founding male's surname and Y lineage had been lost by drift. The eighth extant Y lineage appeared to represent gene flow from outside. The Samaritans are a distinct religious and cultural community in the Middle East who split from mainstream Judaism around 2,500 years ago and numbered several thousand during the Roman period. A genetic survey, also published in 2003, found just four main Y lineages (and close STR variants of some) (Bonné-Tamir et al. 2003). Two of the four lineages shared a common ancestor estimated to date to approximately the time when the population was established, so it seems likely

that all surviving Y lineages trace back to three founders ~2,500 years ago: a striking illustration of genetic drift. A rather larger population, that of Iceland, was established in approximately 870 AD by between 8,000 and 20,000 individuals, and now numbers around 280,000, mostly as a result of endogenous growth since there has been little subsequent immigration. Y-chromosomal diversity is grossly comparable to that of nearby European countries, but enhanced genetic drift is detectable by some measures (Helgason et al. 2003b), and large-scale genealogical studies reveal that the 71% of the contemporary male population whose ancestry can be traced back three hundred years (approximately eight generations) descend from only 10% of the population (Helgason et al. 2003a). We thus see lineage loss in all populations, but most markedly in the Samaritans, whose size, degree of endogamy and timeframe could provide a model for some Indian populations.

DATA SOURCES AND ANALYSES

We sought datasets from India and East Asia that reported both Y-STR and Y-SNP genotypes from reasonably-sized population samples (Table 1). It was necessary to strike a balance between the number of markers and number of populations included. When we set a requirement for a minimum sample size of 17 males typed with 31 Y-SNPs and 9 Y-STRs, we were able to analyse 1,764 individuals: 784 from 31 populations in India (Sengupta et al. 2006; Zerjal et al. 2007) and 980 from 27 populations in East Asia, mainly China (Xue et al. 2006).

Diversity within individual populations or groups of populations was summarised by (i) Nei's gene (= STR haplotype) diversity (Nei 1987), (ii) the average squared distance between haplotypes (ASD), and (iii) the population mutation parameter θ_k (Ewens 1972); if the average mutation rate is similar in different populations, variation in θ_k will reflect variation in the male effective population size. Genetic distance measures between pairs of populations were (i) F_{ST} (for Y-SNPs), (ii) R_{ST} (for Y-STRs; Slatkin 1995), (iii) ASD and (iv) ρ (the distance between a haplotype in one population and the closest haplotype in the second population, averaged over all haplotypes). In comparisons of these measurements, we report the median value rather than the mean, because the measurements were often not normally distributed. Medians were compared using Mann-Whitney *U* tests and in some cases multidimensional scaling (MDS) plots were constructed; both analyses were performed using SPSS 14.0. Analysis of Molecular Variance (AMOVA) was carried out with Arlequin (Schneider et al. 2000).

Y-CHROMOSOMAL VARIATION IN CASTE, TRIBAL AND EAST ASIAN POPULATIONS

Data were available from 19 caste and 12 tribal populations within India and 27 populations from East Asia. In considering these data, we concentrate mainly on Y-STRs because they are less affected by marker ascertainment bias than Y-SNPs; 'variation' thus implies 'STR variation' unless otherwise stated.

Variation within a population can be summarised by several statistics, and we used haplotype diversity, θ_k and ASD (Table 1). Median values of all these measures were lower in tribes than castes, and were lower in both Indian groups than in East Asia, except that ASD was slightly lower in East Asia than in castes (Table 2). A Mann-Whitney *U* test was used to assess the significance of the differences between the caste, tribal and East Asian groups and they were found to be significant in all comparisons, except for the East Asia-caste ASD difference mentioned above (Table 3). The measures used reflect related, but slightly different, features of the population variation. ASD takes into account the molecular differences between haplotypes and it is likely that the caste populations, who have significantly lower haplotype diversity than the East Asians but similar ASD, contain some highly divergent haplogroups and these molecular differences contribute more to the ASD statistic than to the diversity value. Indeed, a single predominant haplogroup, O, was noted in East Asia (Xue et al. 2006), but there was more variety of haplogroups in India (Sengupta et al. 2006; Zerjal et al. 2007). Overall, there is thus a strong and clear pattern of *within*-population variation: tribes<castes<East Asia.

For comparisons of variation *between* populations within the caste, tribal and East Asian groups, we used the measures F_{ST}, R_{ST}, ASD and ρ. We emphasise that all the comparisons are of genetic distances between one caste population and another, between one

Table 1: Population samples included in this work.

Country	*Population name*	*Social category*	*N*	*Number of haplotypes*	*Haplotype diversity*	θ_K	*ASD*	*Reference*
India	Ambalakarar	caste	29	22	0.975	40	97	Sengupta et al. 2006
India	Brahmin_Indians	caste	17	15	0.985	57	92	Zerjal et al. 2007
India	Brahmin_Jaunpur	caste	20	10	0.837	7	43	Zerjal et al. 2007
India	Chamar	caste	17	13	0.963	23	46	Sengupta et al. 2006
India	Iyengar	caste	29	28	0.998	387	95	Sengupta et al. 2006
India	Iyer	caste	29	26	0.993	117	86	Sengupta et al. 2006
India	Koknasth Brahmin	caste	25	20	0.980	44	67	Sengupta et al. 2006
India	Kshatriyas_Indians	caste	19	18	0.994	159	83	Zerjal et al. 2007
India	Kshatriyas_Jaunpur	caste	47	14	0.642	6	14	Zerjal et al. 2007
India	Maratha	caste	20	19	0.995	177	97	Sengupta et al. 2006
India	Other_Indians	caste	23	23	1.000	∞	96	Zerjal et al. 2007
India	Pallan	caste	27	27	1.000	∞	100	Sengupta et al. 2006
India	Panchamas_Indians	caste	19	15	0.959	31	106	Zerjal et al. 2007
India	Panchamas_Jaunpur	caste	28	20	0.968	30	99	Zerjal et al. 2007
India	Rajput	caste	29	28	0.998	387	78	Sengupta et al. 2006
India	Vaishyas_Jaunpur	caste	39	30	0.973	58	77	Zerjal et al. 2007
India	Vanniyar	caste	24	23	0.996	260	82	Sengupta et al. 2006
India	Vellalar	caste	28	14	0.937	10	48	Sengupta et al. 2006
India	West Bengal Brahmin	caste	17	13	0.971	23	49	Sengupta et al. 2006
India	Halba	tribe	20	16	0.974	35	71	Sengupta et al. 2006
India	Ho	tribe	30	18	0.883	18	52	Sengupta et al. 2006
India	Irula	tribe	30	23	0.982	43	61	Sengupta et al. 2006
India	Jamatia	tribe	30	19	0.966	21	63	Sengupta et al. 2006
India	Kamar	tribe	30	10	0.807	5	45	Sengupta et al. 2006
India	Konda Reddy	tribe	29	21	0.968	33	46	Sengupta et al. 2006
India	Koya Dora	tribe	27	21	0.977	42	70	Sengupta et al. 2006
India	Kurumba	tribe	17	10	0.868	9	28	Sengupta et al. 2006
India	Lodha	tribe	20	15	0.963	26	45	Sengupta et al. 2006
India	Mizo	tribe	27	23	0.989	71	56	Sengupta et al. 2006
India	Muria	tribe	18	10	0.843	8	40	Sengupta et al. 2006
India	Tripuri	tribe	20	15	0.968	26	33	Sengupta et al. 2006
China	Buyi	N.A.[a]	35	32	0.995	176	45	Xue et al. 2006
China	Chinese Korean	N.A.	25	23	0.993	134	86	Xue et al. 2006
China	Daur	N.A.	39	29	0.976	50	75	Xue et al. 2006
China	Ewenki	N.A.	26	21	0.979	49	74	Xue et al. 2006
China	Han (Chengdu)	N.A.	34	34	1.000	∞	69	Xue et al. 2006
China	Han (Harbin)	N.A.	35	35	1.000	∞	68	Xue et al. 2006
China	Han (Lanzhou)	N.A.	30	28	0.995	198	99	Xue et al. 2006
China	Han (Meixian)	N.A.	35	35	1.000	∞	57	Xue et al. 2006
China	Han (Yili)	N.A.	32	32	1.000	∞	103	Xue et al. 2006
China	Hani	N.A.	34	29	0.989	90	59	Xue et al. 2006
China	Hezhe	N.A.	44	41	0.997	287	78	Xue et al. 2006
China	Hui	N.A.	35	28	0.983	63	77	Xue et al. 2006
China	Li	N.A.	34	21	0.859	22	32	Xue et al. 2006
China	Manchu	N.A.	35	35	1.000	∞	95	Xue et al. 2006
China	Inner Mongolian	N.A.	45	39	0.993	136	94	Xue et al. 2006
China	Oroqen	N.A.	31	27	0.979	96	53	Xue et al. 2006
China	Qiang	N.A.	33	33	1.000	∞	81	Xue et al. 2006
China	She	N.A.	34	22	0.963	26	43	Xue et al. 2006
China	Tibetans	N.A.	35	31	0.992	126	114	Xue et al. 2006
China	Uygur (Urumqi)	N.A.	39	39	1.000	∞	102	Xue et al. 2006
China	Uygur (Yili)	N.A.	31	31	1.000	∞	135	Xue et al. 2006
China	Xibe	N.A.	41	41	1.000	∞	92	Xue et al. 2006
China	Yao (Bama)	N.A.	35	17	0.908	12	46	Xue et al. 2006
China	Yao (Liannan)	N.A.	35	29	0.988	77	55	Xue et al. 2006
Japan	Japanese	N.A.	47	45	0.998	510	82	Xue et al. 2006
Korea	Korean	N.A.	43	40	0.997	273	78	Xue et al. 2006
Mongolia	Outer Mongolian	N.A.	58	49	0.993	146	79	Xue et al. 2006
	Total		1764					
India	Jaunpur artificial	caste	35	23	0.955	28	95	This work

[a] Not applicable

Table 2: Median values of within-population Y-STR variation statistics

Group of populations	θ_K *diversity*	*ASD*	
Castes	0.980	57	83
Tribes	0.967	26	49
East Asia	0.995	175	77
Local castes*	0.975	44	78
Local Chinese*	0.993	136	74

*See text for the composition of these groups

Table 3: Comparisons of within-population Y-STR variation between caste, tribal and East Asian groups*

Group compared	*Haplotype diversity*	θ_K	*ASD*
Castes v East Asians	0.040	0.024	0.616
Tribes v East Asians	<0.001	<0.001	0.001
Castes v tribes	0.048	0.032	0.002

*Mann-Whitney *U* test P value, 2-sided

tribal population and another, or between one East Asian population and another; none of the distances are between the groups. ASDs were similar within all groups perhaps reflecting the presence of diverse ancient haplogroups within each group. All the other measures, however, differed between the groups and showed a common trend, but one that was different from the within-population trend: castes<East Asia<tribes (Table 4). Apart from ASD, all of these differences were significant (Table 5). The F_{ST} and R_{ST} results are illustrated in the MDS plots (Fig. 1), where the wide scatter of tribal Indian populations is particularly apparent.

While it is unsurprising to find that the

Table 4: Median values of between-population within-group Y-chromosomal variation statistics

Group of populations	F_{ST}	R_{ST}	*ASD*	ρ
Castes	0.058	0.034	89	3.6
Local castes*	0.067	0.069	86	3.5
Tribes	0.260	0.262	81	4.7
East Asia	0.096	0.064	86	3.9
Local Chinese*	0.088	0.060	82	3.8

*See text for the composition of these groups

Table 5: Comparisons of between-population Y-chromosomal variation within caste, tribal and East Asian groups*

Groups compared	F_{ST}	R_{ST}	*ASD*	ρ
Castes v East Asia	<0.001	0.002	0.678	<0.001
Tribes v East Asia	<0.001	<0.001	0.442	<0.001
Castes v tribes	<0.001	<0.001	0.326	<0.001
Local castes v local Chinese	0.465	0.942	0.137	0.002
Tribes v local Chinese	<0.001	<0.001	0.455	<0.001
Local castes v tribes	<0.001	<0.001	0.932	<0.001

*Mann-Whitney *U* test P value, 2-sided

patterns of within-population and between-population variation differ, we might expect that strong genetic drift would lead to both low variation within populations and large differences between populations. According to this simple model, if genetic drift were highest in tribal populations, intermediate in caste populations and lowest in East Asian populations, we would see the observed tribes<castes <East Asians pattern of within-population

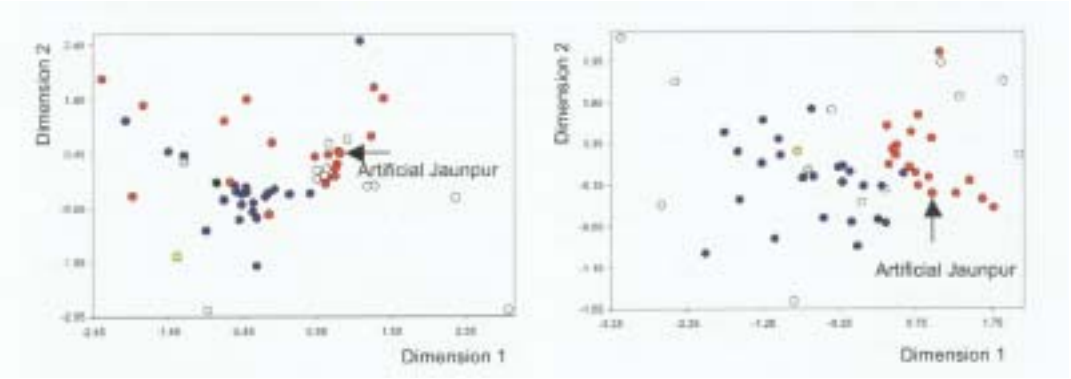

Fig. 1. Genetic distances between populations in India and East Asia. Each circle represents a population sample: red = Indian caste, white = Indian tribe, blue = China, yellow = Korea, grey = Japan and black = Mongolia. A. F_{ST} genetic distances using 31 Y-SNPs, MDS plot (stress = 0.17, R^2 = 0.90). B. R_{ST} genetic distances using 9 Y-STRs, MDS plot (stress = 0.15, R^2 = 0.91).

variation, but would see the converse pattern of East Asians<castes<tribes for between-population variation. We therefore investigated whether the observed between-population variation order of castes<East Asians<tribes, which did not fit this simple expectation, could result from the sampling strategy used. We repeated the analyses using only caste samples from local regions (i.e. excluding all of the mixed 'Indian' samples of Zerjal et al.) and restricting the East Asian samples to populations who have been resident in China for a long time (i.e. excluding the Mongolian, Korean and Japanese populations and also the Uygur and Hui who have entered China within historical times). These changes had no substantial effect on either the relative within-population variation (Table 2, lower section) or the comparison between either group and tribal populations, but did lead to the 'local caste' and 'local Chinese' groups being similar, except with the ρ measure (Tables 4 and 5). This result seems the most reliable one: we therefore conclude that between-population variation follows the order [East Asians/castes]<tribes.

AMOVA analysis allows variation to be apportioned between categories in a quantitative way. We first analysed data from India and China separately, and calculated the percentage of variance within and between populations in each country (Table 6). With both Y-SNPs and Y-STRs, India shows more than twice the amount of variation between populations that is seen in China. With Y-STRs, for example, these results correspond to a F_{ST} of 0.21 in India, compared with 0.08 in China. When the Indian populations were grouped in caste and tribal groups, and compared with the control group from East Asia/China, substantial variation was seen both between populations within a group and between groups. The results were broadly similar for the different markers and group compositions, and always showed more variation in the 'between-population within-group' category than in the 'between-group' category (Table 6).

In summary, a simple pattern of Y-chromosomal variation emerges when Indian populations are compared with East Asian ones: in India, variation *within* populations is lower, and variation *between* populations is, on average, higher. The effect is more marked for the tribal samples analysed here than for the caste samples; indeed the variation between caste populations

Table 6: AMOVA analysis

Grouping	*Marker*	*Proportion of variation (%)*	
		Within populations	*Between populations*
India	Y-SNPs	77.1	22.9
India	Y-STRs	78.6	21.4
China	Y-SNPs	88.9	11.1
China	Y-STRs	92.4	7.6

		Within populations	*Between populations, within groups*	*Between groups*
Castes, tribes, East Asians	Y-SNPs	78.1	13.6	8.3
Castes, tribes, East Asians	Y-STRs	81.2	11.2	7.6
Local castes, tribes, local Chinese	Y-SNPs	75.2	14.2	10.6
Local castes, tribes, local Chinese	Y-STRs	78.5	12.3	9.1

was similar to the variation between Chinese populations. Our results thus emphasise the unusual nature of the genetic structure in India and show that the so-called grandest genetic experiment has had detectable effects in this part of the world.

DISCUSSION

The caste system created social substructure for millennia within the Indian population. If the resulting subpopulations were sufficiently small and genetically isolated, and existed for long enough for genetic drift to be effective, this social substructure would lead to detectable genetic substructure (Fig. 2). Previous work has suggested that the conditions necessary for significant genetic drift were likely to have been met, at least for some populations. A study of caste populations from the Jaunpur district, for example, estimated male effective population sizes as small as 800 and 690 for Brahmins and Kshatriyas, respectively, although estimated sizes for Vaishyas and Panchamas were larger: 2,300 and 2,500 (Zerjal et al. 2007), and may well be different for all castes in other regions. The same study estimated gene flow from the Kshatriyas into all other castes in the same location at approximately 0.7% per generation, similar to the value of 1-2% per generation estimated by other workers (Wooding et al. 2004). In the present broader analysis, we found that individual caste

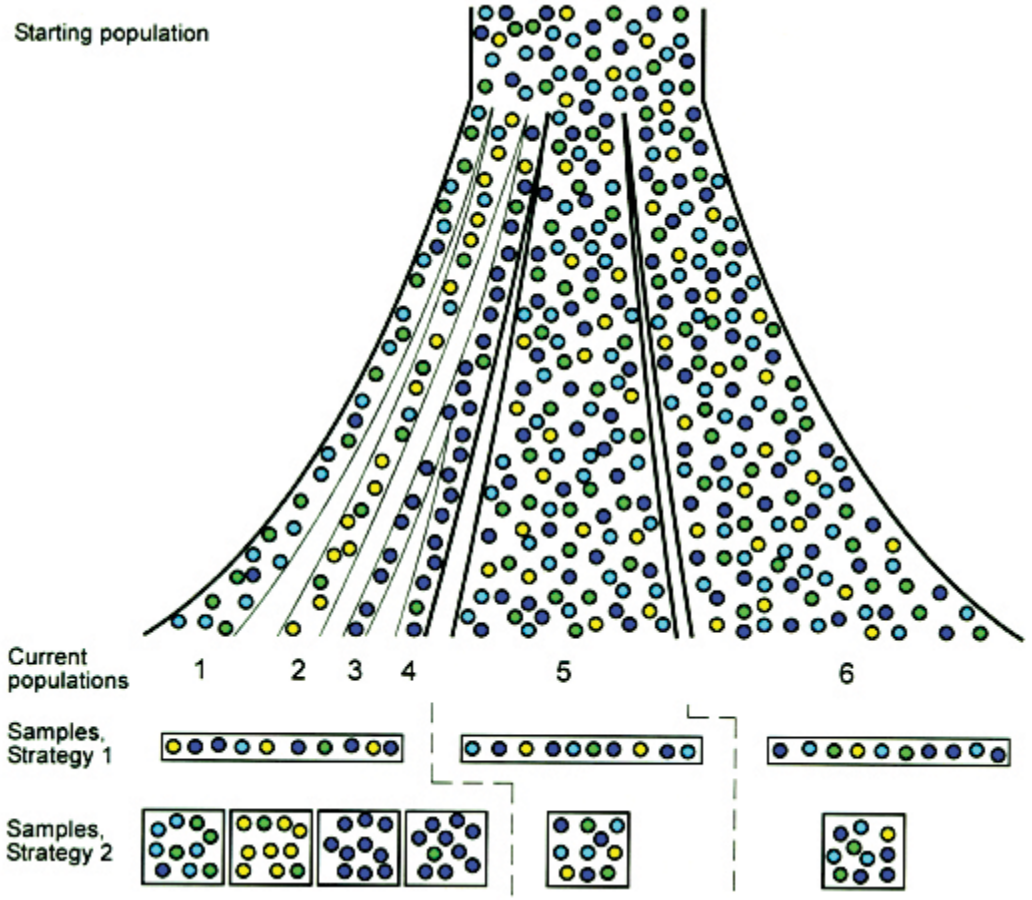

Fig. 2. Effects of subdivision, population size and sampling strategy. Each circle represents a Y-chromosomal lineage and the lines represent boundaries between six populations. In the four smaller populations, there has been substantial genetic drift leading to a single predominant lineage in each. When each of the six populations is sampled separately (Strategy 2) variation within populations 1-4 is low, and variation between them is high, while the converse is seen for populations 5 and 6. However, when a sampling strategy is applied that does not take account of the distinctions between populations 1-4 (Strategy 1), the effects of subdivision are no longer seen.

populations generally contained significantly less variation than East Asian populations as would be expected if they had experienced more genetic drift, but this did not lead to them being more distinct from other caste populations, which would also be expected from a simple model of drift in a subdivided population (Fig.2). Interestingly, the observations of low within-population variation combined with high between-population variation were much more striking in the tribal population samples examined. This could reflect smaller effective population sizes in the tribes, less gene flow, a longer time period of population subdivision or any combination of these factors. However, another factor also needs to be taken into account when considering these results: the sampling strategy.

The criteria for choosing particular samples are often unclear, and may be opportunistic, reflecting the individuals and populations who wished to participate in a study. The sampling strategy adopted in any genetic survey is always very important, but can have a far greater influence on the conclusions in a highly substructured population than in one with low levels of structure. Consider the six hypothetical

current populations illustrated in Figure 2. In sampling strategy 1, the investigators do not take account of the subdivision between populations 1-4, but combine them into a single population and compare them with populations 5 and 6. They conclude that all populations contain high levels of within-population variation and that differences between them are low. In contrast, in sampling strategy 2, investigators sample populations 1-4 separately and consequently detect the low levels of variation within some populations and high levels of variation between populations.

To illustrate the magnitude of this effect in an Indian context, where there is clear geographical structure (e.g. Gutala et al. 2006; Reddy et al. 2005), we re-analyse the published data from individual castes in Jaunpur (Zerjal et al. 2007) by pooling them into a single artificial 'Jaunpur caste' sample of 35, consisting of an arbitrary seven individuals from each of the castes combined into a single pseudo-population. The within-population variation measures of haplotype diversity, θ_K and ASD are no longer exceptionally low (Table 1, last row). The individual Jaunpur castes were very distinct from some other caste populations: for example, the R_{ST} distances between Jaunpur Brahmins, Kshatriyas, Vaishyas and Panchamas and the Vellalar middle caste sample of Sengupta et al. (2006) were 0.289, 0.550, 0.100 and 0.097 respectively. In contrast, the distance between the artificial Jaunpur caste sample and the Vellalar was 0.135. The overall effect on between-population distances can be seen in the MDS plots (Fig. 1): the artificial sample lies well within the cluster of caste populations, while some of the individual castes are extreme outliers. In this illustration, the populations considered were different castes, but the same effects could potentially be seen with tribes or breeding isolates within a caste.

The comparison of Indian with East Asian populations thus reveals several, but not all, of the features expected from a simple increase in genetic drift if the Indian population is more subdivided: variation within populations is lower in India, and variation between tribal populations is higher, as expected, but variation between caste populations is not higher than between East Asian populations. Sampling strategy is rarely described in detail and may have influenced this conclusion, for example if the caste samples do not correspond to true endogamous groups. Sampling procedures should be described in detail. Alternatively, from a Y-chromosomal perspective, the 'grandest experiment ever performed' may in fact have been the one which produced the tribal social and genetic structure, rather than the caste system.

ACKNOWLEDGEMENTS

We thank Mohan Reddy for helpful comments on 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

Bhattacharyya NP, Basu P, Das M, Pramanik S, Banerjee R, Roy B, Roychoudhury S, Majumder PP 1999. Negligible male gene flow across ethnic boundaries in India, revealed by analysis of Y-chromosomal DNA polymorphisms. *Genome Res*, **9:** 711-719.

Bonné-Tamir B, Korostishevsky M, Redd AJ, Pel-Or Y, Kaplan ME, Hammer MF 2003. Maternal and paternal lineages of the Samaritan isolate: mutation rates and time to most recent common male ancestor. *Ann Hum Genet*, **67:** 153-164.

Census of India 2001 (on line) (consulted 04-02-2007). <http://www.censusindia.net/results/population.html>

Dobzhansky T 1973. *Genetic Diversity and Human Equality*. New York: Basic Books.

Ewens WJ 1972. The sampling theory of selectively neutral alleles. *Theor Popul Biol*, **3:** 87-112.

Gadgil M, Malhotra KC 1983. Adaptive significance of the Indian caste system: an ecological perspective. *Ann Hum Biol*, **10:** 465-477.

Gutala R, Carvalho-Silva DR, Jin L, Yngvadottir B, Avadhanula V, Nanne K, Singh L, Chakraborty R, Tyler-Smith C 2006. A shared Y-chromosomal heritage between Muslims and Hindus in India. *Hum Genet*, **120:** 543-551.

Helgason A, Hrafnkelsson B, Gulcher JR, Ward R, Stefánsson K 2003a. A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. *Am J Hum Genet*, **72:** 1370-1388.

Helgason A, Nicholson G, Stefánsson K, Donnelly P 2003b. A reassessment of genetic diversity in Icelanders: strong evidence from multiple loci for relative homogeneity caused by genetic drift. *Ann Hum Genet*, **67:** 281-297.

Jobling MA, Tyler-Smith C 2000. New uses for new haplotypes: the human Y chromosome, disease and selection. *Trends Genet*, **16:** 356-362.

Jobling MA, Tyler-Smith C 2003. The human Y chromosome: an evolutionary marker comes of age. *Nature Reviews Genetics*, **4:** 598-612.

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

Nei M 1987. *Molecular Evolutionary Genetics*. New York: Colombia University Press.

Reddy BM, Naidu VM, Madhavi VK, Thangaraj LK, Kumar V, Langstieh BT, Venkatramana P, Reddy AG, Singh L 2005. Microsatellite diversity in Andhra Pradesh, India: genetic stratification versus social stratification. *Hum Biol*, **77:** 803-823.

Schneider S, Roessli D, Excoffier L (2000). Arelquin: a software for population genetics data analysis, Genetics and Biometry Lab., Dept. of Anthropology, University of Geneva.

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 2006. 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*, **78:** 202-221.

Singh KS 1993. *People of India, Volume XI: An Anthropological Atlas*. Delhi: Oxford University Press.

Slatkin M 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139:** 457-462.

Soodyall H, Nebel A, Morar B, Jenkins T 2003. Genealogy and genes: tracing the founding fathers of Tristan da Cunha. *Eur J Hum Genet*, **11:** 705-709.

Thapar R 1990. *A History of India, Volume One*. London: Penguin Books.

Wikipedia: Tristan da Cunha 2007 (on line) (consulted 20.02.2007). http://en.wikipedia.org/wiki/Tristan_da_Cunha

Wolpert S 1997. *A New History of India, Fifth Edition*. Oxford: Oxford University Press.

Wooding S, Ostler C, Prasad BV, Watkins WS, Sung S, Bamshad M, Jorde LB 2004. Directional migration in the Hindu castes: inferences from mitochondrial, autosomal and Y-chromosomal data. *Hum Genet*, **115:** 221-229.

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 2006. Male demography in East Asia: a north-south contrast in human population expansion times. *Genetics*, **172:** 2431-2439.

Zerjal T, Pandya A, Thangaraj K, Ling EY, Kearley J, Bertoneri S, Paracchini S, Singh L, Tyler-Smith C 2007. Y-chromosomal insights into the genetic impact of the caste system in India. *Hum Genet*, **121:** 137-144.