

The Northeast Indian Passageway: A Barrier or Corridor for Human Migrations?

Richard Cordaux,*† Gunter Weiss,‡ Nilmani Saha,§ and Mark Stoneking*

*Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; †Department of Biological Sciences, Louisiana State University, Baton Rouge; ‡WE Informatik, Heinrich-Heine-Universität Duesseldorf, Duesseldorf, Germany; and §Department of Pediatrics, National University of Singapore, Singapore

The northeast Indian passageway connecting the Indian subcontinent to East/Southeast Asia is thought to have been a major corridor for human migrations. Because it is also an important linguistic contact zone, it is predicted that northeast India has witnessed extensive population interactions, thus, leading to high genetic diversity within groups and heterogeneity among groups. To test this prediction, we analyzed 14 biallelic and five short tandem-repeat Y-chromosome markers and hypervariable region 1 mtDNA sequence variation in 192 northeast Indians. We find that both northeast Indian Y chromosomes and mtDNAs consistently show strikingly high homogeneity among groups and strong affinities to East Asian groups. We detect virtually no Y-chromosome and mtDNA admixture between northeast and other Indian groups. Northeast Indian groups are also characterized by a greatly reduced Y-chromosome diversity, which contrasts with extensive mtDNA diversity. This is best explained by a male founder effect during the colonization of northeast India that is estimated to have occurred within the past 4,000 years. Thus, contrary to the prediction, these results provide strong evidence for a genetic discontinuity between northeast Indian groups and other Indian groups. We, therefore, conclude that the northeast Indian passageway acted as a geographic barrier rather than as a corridor for human migrations between the Indian subcontinent and East/Southeast Asia, at least within the past millennia and possibly for several tens of thousand years, as suggested by the overall distinctiveness of the Indian and East Asian Y chromosome and mtDNA gene pools.

Introduction

Modern human evolution is punctuated with migrations that ultimately led our ancestors to spread over most parts of the world within the past 100,000 years (Cavalli-Sforza, Menozzi and Piazza 1994; Lahr and Foley 1994). Migration routes are influenced by landscape and, consequently, natural passageways acting as corridors played a key-role in the history of human migrations, as exemplified by the Nile River Valley, which connects Africa and Eurasia (Krings et al. 1999), central Asia, which connects west and east Eurasia (Karafet et al. 2001; Wells et al. 2001; Zerjal et al. 2002), and eastern Indonesia, which connects Southeast Asia and Australia/Oceania (Kayser et al. 2000a).

The northeastern tip of India, flanked in the north by the Himalayas and in the south by the Bay of Bengal, constitutes a unique narrow passageway that connects the Indian subcontinent to East Asia and Southeast Asia (fig. 1). It is thought to have been a crucial corridor for human migrations between these two subcontinental areas, including, perhaps, the first migrations from Africa towards East Asia and Australia more than 40,000 years ago (Nei and Roychoudhury 1993; Cavalli-Sforza, Menozzi, and Piazza 1994; Lahr and Foley 1994; Cann 2001).

Northeast India is also an area of extensive linguistic diversity where three language families are represented: Austro-Asiatic, Indo-European and Tibeto-Burman (Masica 1991; Ruhlen 1991). Austro-Asiatic languages are now spoken by a single group in northeast India (the Khasi, composed of many subgroups), but these languages are also found in east India and Southeast Asia (Ruhlen

1991), which suggests that they might have been more widespread in northeast India in the past (Abbi 1991). Indo-European languages are nowadays spoken from Europe to central and south Asia; the easternmost occurrences are in Nepal, Bangladesh, and northeast India. The arrival of Indo-European speakers into India within the past 3,500 years had a substantial impact on Indian genetic diversity (Bamshad et al. 2001; Quintana-Murci et al. 2001; Wells et al. 2001; Basu et al. 2003; Cordaux et al. 2004). Tibeto-Burman languages are a branch of the Sino-Tibetan family, which is mainly spoken in northeast India, China, and Southeast Asia (Ruhlen 1991). Thus, northeast India constitutes an important linguistic contact zone.

A putative long history of migrations, coupled with diverse cultural influences as evidenced by the high linguistic diversity, would predict that northeast India has experienced extensive population interactions that have resulted in high genetic diversity within groups and heterogeneity among groups. To test this hypothesis, we analyzed Y-chromosome and mitochondrial DNA (mtDNA) variation in the Adi, Apatani, Nishi, and Naga tribal populations, sampled from diverse northeast Indian localities. They speak Tibeto-Burman languages and are also conversant in Indo-European languages, such as Hindi or Assamese (Singh 1998). Genetic variation in these groups was compared with that of other Indian and East/Southeast Asian groups. We find that, contrary to the prediction, northeast Indian groups show a striking genetic homogeneity both in terms of Y-chromosome and mtDNA variation, which was probably maintained over time by genetic isolation. In addition, northeast Indians show virtually no genetic admixture with other Indian groups, which has led to a remarkable genetic discontinuity between these groups. This suggests that the northeast Indian passageway was a geographic barrier to contacts between the Indian subcontinent and East/Southeast Asia, rather than a corridor, at least within the past millennia.

Key words: northeast India, humans, genetic diversity, Y chromosome, mtDNA.

E-mail: cordaux@lsu.edu.

Mol. Biol. Evol. 21(8):1525–1533. 2004

doi:10.1093/molbev/msh151

Advance Access publication May 5, 2004

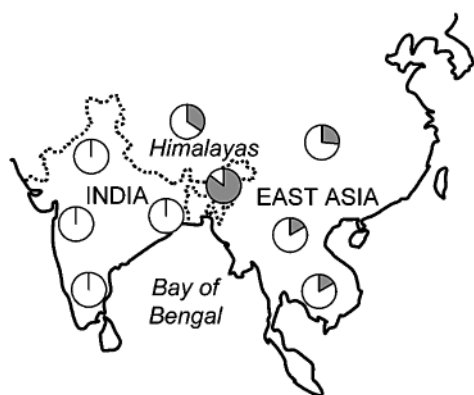


FIG. 1.—Map of India and East/Southeast Asia that indicates the frequency distribution of Y-haplogroup O-M134 (in black).

Subjects and Methods

Subjects

DNA samples from 173 males belonging to four northeast Indian tribal populations from Assam and Arunachal Pradesh were analyzed. The analysis of the samples was carried out with the informed consent of the donors. Additional information on these groups can be found elsewhere (Clark et al. 2000; Cordaux et al. 2003). The mtDNA hypervariable region 1 (HV1) was also analyzed in 192 individuals from the same groups (Cordaux et al. 2003). Available data from three other northeast Indian Tibeto-Burman groups, consisting of Y-chromosome data from 20 Kachari and 17 Rabha (Su et al. 2000b) and mtDNA data from 20 Tipperah (Roychoudhury et al. 2001), were included in some analyses.

Y Chromosome Typing

Fourteen slowly evolving Y-chromosome biallelic polymorphisms were typed. Markers M9, M17, M20, M52, M74, M89, M122, M124, M172, M175, RPS4Y, and YAP were typed as described (Hammer and Horai 1995; Kayser et al. 2000a, 2001, 2003; Ke et al. 2001; Cordaux et al. 2004). Markers M134 and M174 were typed by PCR-RFLP using the procedure described by Cordaux et al. (2004); primers and PCR conditions for these two markers are given in table 1. The nomenclature used is that of the Y-Chromosome Consortium 2003 (Jobling and Tyler-Smith 2003).

In addition, six Y-chromosome short tandem-repeat (Y-STR) loci (DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393) were assayed in 139 northeast Indian samples belonging to Y-haplogroup O-M134, as described previously (Kayser et al. 1997). However, for DYS389, the larger product conventionally ascribed to DYS389II could not reliably be amplified in all samples and was, therefore, excluded from the analyses.

Statistical Analyses

The software package ARLEQUIN version 2.0 (Schneider, Roessli, and Excoffier 2000) was used to calculate Y-haplogroup diversity and F_{st} distances between

pairs of populations and associated P values based on 1,000 permutations. The Mann-Whitney U test to compare diversity values was computed with STATISTICA. χ^2 tests to compare haplogroup frequency distributions was calculated in EXCEL (Microsoft). Analyses of molecular variance (AMOVA) were performed by use of ARLEQUIN, and the significance of variance components were tested with 10,000 permutations. Multidimensional scaling (MDS) analysis was performed by means of STATISTICA, based on F_{st} distances. Populations from northeast India were compared with 131 tribal and 24 caste south Indians (Cordaux et al. 2004), 72 west Indians, 66 north Indians, 31 east Indians (Kivisild et al. 2003), 46 Tibetans, 365 Han Chinese, 76 south Chinese (Su et al. 2000b), and 71 Southeast Asians (Su et al. 2000a). The Indian and East/Southeast Asian contributions to the northeast Indian Y-chromosome gene pool were estimated by use of (1) a phylogeographic approach (Cordaux et al. 2004) and (2) the program ADMIX version 2.0 (Dupanloup and Bertorelle 2001). The two parental populations were represented by the 324 Indian and 558 East/Southeast Asian individuals described above, and standard errors (SE) were calculated on the basis of 1,000 bootstraps.

The Y-STRs were recorded as haplotypes. ARLEQUIN was used to calculate (1) Y-STR haplotype diversity, and (2) mean pairwise differences (MPD), the mean number of mutational steps observed between all pairs of haplotypes in the sample. Variance in allele size distribution was calculated in EXCEL for each locus independently and then averaged across the five loci. A median-joining network connecting the different Y-STR haplotypes was constructed by utilizing the NETWORK version 3.1 software (Bandelt, Forster, and Rohl 1999). Locus-specific weights were given according to Kayser et al. (2000a, 2000b), so that loci with the highest mutation rates were given the lowest weights. Hence, DYS389I, DYS390, DYS391, DYS392, and DYS393 were given weights of 5, 1, 2, 10, and 10, respectively. In addition, Y-STR haplotypes were compared with 109 haplotypes from East, Southeast, and island Southeast Asia (Kayser et al. 2000a, 2003) to estimate the extent of haplotype sharing between these groups and northeast Indian groups.

A coalescence analysis of the 139 Y-STR haplotypes was performed by use of the BATWING version 1.0 software (Wilson, Weale, and Balding 2003). A two-phase population model was chosen, in which population size in the past was constant, and then experienced a period of exponential growth until the present. Thus, the demography of the population is defined by three parameters: initial population size, growth rate, and time since expansion. Details on the prior distribution characteristics used to model the three aforementioned demographic parameters are given in Kayser et al. (2000a, 2001). In brief, they cover a range of demographic scenarios that range from no growth (constant population size) to reasonable growth rates for human populations (Kayser et al. 2000a, 2001). Gamma-distributed prior distributions were assigned to the mutation rates of the five STR loci, adjusted to the corresponding estimates reported in Kayser et al. (2000b). A Markov chain Monte Carlo method was then used to generate approximate random samples for converting the

Table 1
PCR-RFLP Parameters for Y-Chromosome Biallelic Polymorphisms

Marker	Y-Hg ^a	PCR Primers (5'→3')		T _{ann} ^b (°C)	RFLP Enzyme	RFLP Fragments (bp)	
		Forward	Reverse			Ancestral	Derived
M174 (T→C)	D-M174	tcttctcgtcacagcaaaaatg	gcaaatgcaccctcacttct	56	<i>Bfal</i>	96+27	123
M134 (G→del)	O-M134	aatcatcaaacaccagaagggt	gatacttttgatccccacgaa	54	<i>TfiI</i>	53+21	73

^a Y-chromosome haplogroup name.^b Annealing temperature.

prior distributions into posterior distributions, which, in turn, reflect the information contained in the data.

mtDNA Analyses

The ADMIX software was used to evaluate the Indian and East/Southeast Asian contributions to the northeast Indian mtDNA gene pool. The Indian parental population was represented by 560 tribal and nontribal individuals (Cordaux et al. 2003). The East/Southeast Asian parental population was represented by 742 individuals (see Cordaux et al. [2003] and references therein). Standard errors were estimated on the basis of 1,000 bootstraps. Fst distances based on mtDNA HV1 sequences were calculated in ARLEQUIN.

The 192 northeast Indian mtDNA sequences were also used to estimate expansion time τ (in mutational units), as implemented in ARLEQUIN. The expansion time t (in years) was then deduced using the relationship $t = \tau / 2u$, where u is the mutation rate for the whole sequence. We used a mutation rate of 1.65×10^{-7} /site/year (Ward et al. 1991) and estimated 95% confidence intervals on the basis of 1,000 bootstraps.

Results

Y-Haplogroup Variation

A total of 11 Y-haplogroups were observed in 173 northeast Indians (fig. 2). However, eight haplogroups were represented by three or fewer individuals, and the single haplogroup O-M134 accounted for 85.5% of all Y chromosomes. This high frequency of O-M134 is characteristic of all northeast Indian groups, because the frequency of O-M134 was not significantly different among populations (range: 76% to 94%; $\chi^2 = 5.7$, $df = 3$, $P = 0.13$). Moreover, O-M134 has previously been reported at frequencies of 85% and 76% in two other Tibeto-Burman northeast Indian groups (Kachari and Rabha [Su et al. 2000b]); these frequencies do not differ from the present results ($\chi^2 = 6.4$, $df = 5$, $P = 0.27$).

Consistent with the high frequency of the single haplogroup O-M134, Y-haplogroup diversity was low in all populations (fig. 2) (average 0.27 [range: 0.11 to 0.39]) and significantly lower (Mann-Whitney U test: $Z = 2.70$, $P < 0.01$) than in 15 south Indian tribal populations (Ramana et al. 2001; Kivisild et al. 2003; Cordaux et al. 2004). The homogeneity in Y-haplogroup composition among northeast Indian groups was further supported by average Fst distances separating the four populations, because the average Fst distance among the four groups was only 0.039, and five out of six pairwise comparisons

yielded Fst values that were not significantly different from zero (table 2). Furthermore, AMOVA results indicated that 96.3% of the variance was within populations, whereas only 3.7% ($P < 0.01$) of the variance was among populations. The latter value contrasts with the 20.5% proportion of variance found among 12 south Indian tribal groups (Cordaux et al. 2004). When the analyses were repeated with Kachari and Rabha included, the average Fst distance between the six northeast Indian groups was only 0.028, and only 2.9% of the variance was among populations.

The geographic distribution of Y-haplogroup O-M134 in eastern Eurasia (fig. 1) suggests an East/Southeast Asian rather than Indian origin of most northeast Indian Y lineages, because O-M134 is typically found in East/Southeast Asia at frequencies up to 25% to 35% and is absent from India and elsewhere (Su et al. 2000a, 2000b; Kivisild et al. 2003; Cordaux et al. 2004). The phylogeography of the Y-haplogroups found in northeast Indian tribal groups suggests that (1) haplogroups D-M174 and O-M175 and its derivatives are of East/Southeast Asian origin (Underhill et al. 2001), (2) haplogroups H-M52 and F-M89 are of indigenous Indian origin (Cordaux et al. 2004), and (3) haplogroups J-M172, L-M20, R-M17, and R-M124 are associated with Indo-European speakers (Cordaux et al. 2004). This scheme allows all but five Y chromosomes to be assigned an origin. Hence, among the 168 assigned individuals, 94% are estimated to have an East/Southeast Asian origin, whereas only 3% are related to Indo-European speakers and 3% are related to indigenous Indian tribal groups. These results are confirmed by an ADMIX

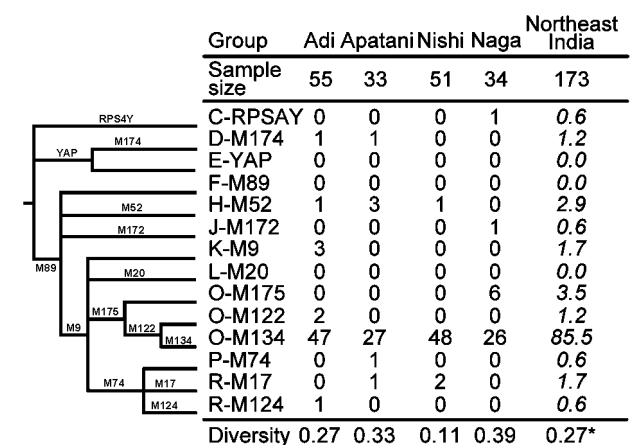


FIG. 2.—Y-chromosome haplogroup frequencies in four northeast Indian tribal populations. Haplogroup relationships are shown with haplogroup-defining markers along the relevant branches of the tree. Italicized values are given in percentages; other values are absolute numbers. Asterisks (*) indicate average value.

Table 2
Average *F_{st}* Values Among Four Northeast Indian Tribal Groups, Based on Y-Haplogroup Frequencies (Below Diagonal) and Mtdna Data (Above Diagonal)

	Adi	Apatani	Nishi	Naga
Adi	—	0.030*	0.020	0.046*
Apatani	-0.002	—	0.014	0.036*
Nishi	0.017	0.030	—	0.028*
Naga	0.047	0.033	0.108*	—

NOTE.—Asterisk (*) indicates significant value after Bonferroni correction for multiple tests.

analysis, because virtually all northeast Indian Y chromosomes are estimated to have East/Southeast Asian origins (table 3). Moreover, the MDS analysis based on *F_{st}* distances indicates closer affinities of northeast Indian groups with East/Southeast Asian groups than with other Indian groups (fig. 3). The fact that northeast Indian groups are not located between other Indian groups and East/Southeast Asian groups in the MDS plot further suggests that they are not admixed, but rather essentially descended from the latter groups.

Y-STR Variation

The much higher frequency of Y-haplogroup O-M134 in northeast Indian groups (~85%) as compared with their closest East/Southeast Asian relatives (~30%) suggests that a founder effect took place during the colonization of northeast India or that males have experienced bottlenecks since the colonization of northeast India. To obtain further insight into the recent demographic history of these groups, five Y-STRs were analyzed in 139 individuals whose Y chromosomes belong to haplogroup O-M134 (table 4). Only 23 distinct haplotypes were identified, four of which (H4, H8, H12, H14) accounted for 80% of all haplotypes. A comparison of the 23 northeast Indian haplotypes to 109 East/Southeast Asian individuals (Kayser et al. 2000a, 2003) shows that nine northeast Indian haplotypes are shared with 13 Han Chinese, seven island Southeast Asians, and one Vietnamese. The most frequent matches of East/Southeast Asian haplotypes with northeast Indian haplotypes involve haplotypes H4 and H12 (five and seven matches, respectively), which are two of the four most frequent haplotypes in northeast India. In total, the nine

Table 3
Results of Admixture Analyses

	Indian Contribution (\pm SE)	East/Southeast Asian Contribution (\pm SE)
Y-Chromosome		
MD+	0.0% (\pm 5.7%)	100.0% (\pm 5.7%)
MD-	0.0% (\pm 5.5%)	100.0% (\pm 5.5%)
mtDNA		
MD+	5.0% (\pm 15.0%)	95.0% (\pm 15.0%)
MD-	13.9% (\pm 4.2%)	86.1% (\pm 4.2%)

NOTE.—Admixture proportions were estimated by taking into account both haplotype frequencies and molecular distances between haplotypes (MD+) or only haplotype frequencies (MD-).

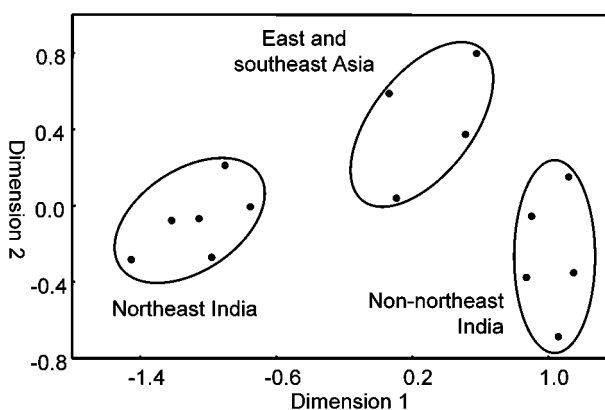


FIG. 3.—MDS plot of 15 Indian and East/Southeast Asian groups, based on Y-haplogroup *F_{st}* distances. Stress value = 0.075.

northeast Indian haplotypes that match East/Southeast Asian haplotypes represent 80 individuals, or 58% of all northeast Indian individuals analyzed here. Thus, there is extensive Y-STR haplotype sharing between northeast India and East/Southeast Asia.

These observations are also reflected in the Y-STR diversity parameters (table 5). There is a reduced number of haplotypes in each population, and accordingly, Y-STR diversity is low (0.66 to 0.75). Moreover, there are, on average, 1.2 differences between any pair of haplotypes drawn from the same population (MPD), and the variance in allele size distribution is only 0.15, which indicates that the different haplotypes are closely related to each other. The close relatedness of the haplotypes is also emphasized in a median-joining network that connects the 23 haplotypes (fig. 4). Indeed, all but two haplotypes can be connected to other haplotypes by a single mutation. In addition, the four most-frequent haplotypes occupy central positions in the network, and most singleton haplotypes occupy more peripheral positions.

A coalescence analysis was performed to gain further insights into the demographic history of northeast Indian groups (table 6 and fig. 5). Growth rates of approximately 0.005, 0.015, and 0.025 per generation indicate slight, moderate, and strong signals of population growth, respectively (Kayser et al. 2000a, 2001). Thus, the growth rate of 0.027/generation detected in northeast India suggests a strong demographic expansion. The analysis also suggests that this expansion most likely started approximately 1,400 years ago from an initial male population size of about 120 individuals. This indicates that the founder effect or bottleneck experienced by northeast Indian males was followed by a strong demographic expansion that started from a reduced number of individuals.

Y-chromosome Versus mtDNA Variation

By contrast to the Y chromosome, which shows greatly reduced diversity within populations (fig. 2 and table 5), mtDNA of northeast Indian groups is among the most diverse on the Indian subcontinent (Cordaux et al. 2003). However, similar to the Y chromosome, mtDNA of northeast Indian groups shows limited differentiation among populations. This is reflected (1) in the small

Table 4
Y-STR Haplotypes Occurring on the Background of Y-Haplogroup O-M134 in 139 Northeast Indian Individuals

Haplotypes	DYS389I	DYS390	DYS391	DYS392	DYS393	Adi	Nishi	Apatani	Naga	Total
H1	9	22	11	14	12	1				1
H2	9	23	9	15	12		1			1
H3	9	23	10	13	12		1			1
H4	9	23	10	14	12	17	18	5	11	51
H5	9	23	10	14	13	2	1			3
H6	9	23	10	15	12	1	3			4
H7	9	23	11	12	12		1			1
H8	9	23	11	14	12	3	3	14		20
H9	9	23	11	14	13				1	1
H10	9	23	12	14	12			1		1
H11	9	24	10	11	12	1				1
H12	9	24	10	14	12	15			4	19
H13	9	24	10	14	13				1	1
H14	9	24	10	15	12	2	15		4	21
H15	9	24	10	15	13			1		1
H16	9	24	10	15	14		1			1
H17	9	24	11	14	12	1	1			2
H18	9	24	11	15	12	1				1
H19	9	25	10	14	12	1				1
H20	9	25	10	15	12		2			2
H21	10	23	10	14	12			1	1	2
H22	10	23	11	14	12			2		2
H23	10	24	10	15	12			1		1
TOTAL						45	47	25	22	139

average F_{st} distance that separates the four groups of the present study (0.029) (table 2), which remains stable when the Tipperah are included (0.030), and (2) in the AMOVA results, which indicate that only 2.9% of the variance is among populations (Cordaux et al. 2003).

With regard to mtDNA relationships, Cordaux et al. (2003) have shown that northeast Indian groups show closer affinities to East/Southeast Asians and are well differentiated from other nonnortheast Indian groups. An ADMIX analysis formally evaluating the East/Southeast Asian and Indian contributions to the northeast Indian mtDNA gene pool confirms this trend, in that East/Southeast Asians have largely (~90%) (table 3) contributed present-day northeast Indian mtDNAs. The mtDNA results, thus, strikingly parallel the Y-chromosome data.

We estimated the expansion time τ to be 6.5 mutational units (95% confidence interval: 4.2 to 7.8), which yielded an expansion time t of 54,000 years (95% confidence interval: 35,000 to 64,000), based on 192 northeast Indian mtDNA sequences. This is in sharp contrast with the expansion time of approximately 1,400 years obtained from Y-STR data (table 6, fig. 5).

Discussion

Y-Chromosome Variation in Northeast India

Tibeto-Burman northeast Indian Y-lineages appear to be more closely related to East/Southeast Asian than to other Indian Y-lineages. This is reflected in (1) the high frequency of haplogroup O-M134, (2) the admixture analyses that suggest almost all northeast Indian Y-chromosomes are of East/Southeast Asian origin, (3) the MDS plot, and (4) the extensive sharing of Y-STR haplotypes between northeast Indians and East/Southeast Asians, particularly Han Chinese.

The demographic scenario suggested by Y-STR variation involves a founder effect; that is, present-day northeast Indian males are derived from a small number of migrants from an East Asian source population. Alternatively, they may have gone through bottlenecks after they reached India. This is supported by (1) the low Y-chromosome diversity (both at haplogroup and STR levels), (2) the fact that only four closely related Y-STR haplotypes encompass 80% of all haplotypes, and (3) all haplotypes are closely related to the four major haplotypes. Moreover, the single Y-haplogroup O-M134 accounts for 85% of all Y-lineages in northeast India, whereas the frequency of O-M134 in the putative source population is approximately 30%. The fact that all northeast Indian populations analyzed consistently show these trends points to a founder effect during the colonization of northeast India rather than to bottlenecks. This is because bottlenecks subsequent to the colonization of northeast India would not be expected to affect all populations in the same way, unless a bottleneck happened right after the founding population arrived in northeast India, before the separation of the different groups. However, a bottleneck immediately

Table 5
Y-STR Diversity Associated with Y-Haplogroup O-M134 in Northeast India

Population	Sample Size	Number of Haplotypes	Diversity (SE)	MPD ^a (SE)	Variance ^b
Adi	45	11	0.71 (0.04)	1.1 (0.7)	0.16
Apatani	25	7	0.66 (0.09)	1.2 (0.8)	0.13
Nishi	47	11	0.75 (0.04)	1.4 (0.8)	0.20
Naga	22	6	0.71 (0.08)	1.2 (0.8)	0.12
Northeast India	139	23	0.71*	1.2*	0.15*

NOTE.—Asterisk (*) indicates average value.

^a Mean pairwise differences.

^b Variance in allele size distribution.

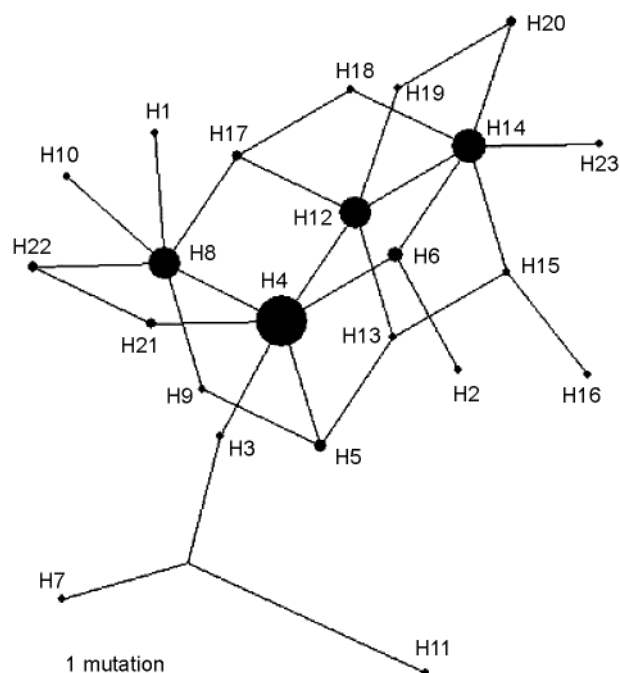


FIG. 4.—Median-joining network of northeast Indian Y-STR haplotypes belonging to haplogroup O-M134. Circles denote haplotypes (listed as in table 4). The size of circles is proportional to the number of individuals carrying this haplotypes. Lines denote Y-STR mutation steps, with a one-step distance being indicated in the lower-right corner.

after arriving in northeast India would essentially be the same as a founder event, for all practical purposes, because it means the reduction in diversity happened either when proto-Tibeto-Burman speakers left east Asia or right when they got to northeast India, not that they were in northeast India for a long time before going through bottlenecks.

This conclusion has important implications because, in the case of separate bottlenecks occurring in the northeast Indian populations, the start of the demographic expansion would provide an indication on the timing of the bottlenecks, not on the timing of the migration to northeast India. However, in the case of a founder effect associated with the colonization of northeast India, the demographic expansion took place immediately or shortly after the colonization, and, hence, the start of the demographic expansion probably represents a good approximation of the time of the migration to northeast India. Our coalescence analysis suggests that the expansion of Tibeto-Burman speakers to northeast India most likely took place within the past 4,200 years, which corresponds to the upper limit of the 95% probability interval of the time since expansion (table 6 and fig. 5B). A fairly recent separation of northeast Indians from their East Asian source is further supported by the low

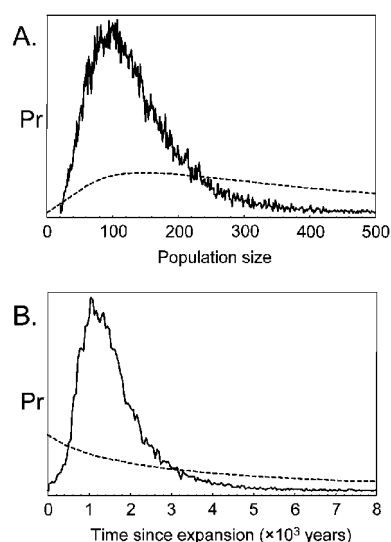


FIG. 5.—Results of the coalescence analysis inferred from Y-STR variation. Each panel shows the prior (dashed lines) and posterior probability distribution of the population size (A) and time since expansion (B). Pr indicates probability.

differentiation of the different northeast Indian groups based on Y-haplogroup frequencies and the extensive sharing of Y-STR haplotypes with East Asians.

Other Y-chromosome evidence, based on 607 individuals speaking Sino-Tibetan languages (to which the Tibeto-Burman family belongs), suggested that the cradle of Sino-Tibetan speakers was in China, perhaps in the Yellow River basin (Su et al. 2000b). Based on the allele size variance of three Y-STRs, Su et al. (2000b) inferred that proto-Tibeto-Burman speakers left China 5,000 to 6,000 years ago, which is consistent with archaeological and linguistic evidence (Wu and Poirier 1995; Etlar 1996) and with the present study. Indeed, almost all northeast Indian Tibeto-Burman Y chromosomes can be assigned an East/Southeast Asian origin, and the admixture with other Indian groups is negligible. This suggests that Tibeto-Burman speakers may have been the first settlers of this area, which is plausible, given the inhospitable topological, climatic, and environmental conditions of the region. Alternatively, Tibeto-Burman newcomers may have found the land already inhabited, in which case they largely replaced the previous inhabitants of northeast India. Possible pre-Tibeto-Burman inhabitants of northeast India are Austro-Asiatic speakers, who nowadays live both west and east of northeast India (Ruhlen 1991) but are represented in northeast India by a single group (the Khasi in Meghalaya [Singh 1998]). The archaeological record of northeast India provides little evidence for pre-Neolithic settlements of the area (Misra 2001), which could be

Table 6
Demographic Inferences Based on Y-STR Data

	Median (95% Probability Interval)		
	Initial Effective Population Size (Individuals)	Population Growth Rate per Generation ($\times 10^{-3}$)	Time Since Expansion (years)
Prior distribution	403 (60–2,860)	6.9 (0.3–36.9)	4,900 (100–64,600)
Posterior distribution	118 (41–326)	27.2 (2.2–80.4)	1,370 (434–4,240)

interpreted as favoring the former hypothesis. However, the northeast Indian archaeological record is very poor, which might explain the paucity of pre-Neolithic evidence.

In addition, the Indo-European component of Tibeto-Burman Y chromosomes is remarkably low (3%), given that although the study groups are primarily Tibeto-Burman speakers, they are also conversant with Indo-European languages. This implies that the “Indo-Europeanization” of northeast India, which is an ongoing process (Masica 1991), is mainly a cultural process. This situation contrasts with southern India, where Indo-European speakers were integrated in non-Indo-European speech communities (Masica 1991) and where Y-chromosome markers typical of Indo-European speakers have been detected in Dravidian-speaking tribal groups (Bamshad et al. 2001; Ramana et al. 2001; Cordaux et al. 2004).

mtDNA Variation in Northeast India

Y-chromosome variation is paternally-inherited and, thus, only reflects male genetic history. How does it compare with mtDNA variation, the female equivalent of the Y chromosome? Y-chromosome and mtDNA variation differ in northeast Indian groups in that reduced Y-chromosome diversity contrasts with high mtDNA diversity, as opposed to other Indian tribal groups (Cordaux et al. 2003; present study). This may be attributable at least in part to the patrilocality of these groups (in which women move to their husband’s residence after marriage), because this leads to diverse Y chromosomes entering a population at a lower rate than mtDNA (Seielstad, Minch, and Cavalli-Sforza 1998; Oota et al. 2001). However, under this scenario, one expects to find a correlation between Y chromosomes and geographic distances and no correlation between mtDNA and geographic distances (Seielstad, Minch, Cavalli-Sforza 1998). Nevertheless, northeast Indian groups do not adhere to this trend (mtDNA: $r^2 = 0.77$, $P = 0.55$; Y chromosome: $r^2 = 0.03$, $P = 0.44$), perhaps because of their low Y-chromosome and mtDNA differentiation or because of the relatively low number of populations compared. Another factor that may have contributed to elevation of mtDNA but not Y-chromosome diversity is gene flow with Indian groups, because mtDNA does indicate about 10% genetic contribution from India (table 3). Alternatively, the different patterns of Y-chromosome and mtDNA diversity observed in northeast India may be best explained by a male-specific founder effect during the colonization of northeast India. This scenario actually finds support in the fact that mtDNA evidence suggests an expansion time approximately 50,000 years ago. This estimate is very similar to that deduced for most human populations (Excoffier and Schneider 1999). Thus, it is more parsimonious to infer that the mtDNA expansion time relates to the human expansion out of Africa, rather than to multiple independent and simultaneous expansions during the colonization of each geographic area. The discrepancy between mtDNA and Y-STR expansion times can be explained by the different mutation rates for each system, which allows them to capture different demographic events during human evolution. It is noteworthy that Excoffier and Schneider (1999) found

that a recent loss of diversity (through bottleneck or founder effect) can erase signals of past population expansion. As the northeast Indian mtDNA gene pool carries a signal of expansion approximately 50,000 years ago (as do the mtDNA gene pool of most other populations), it is reasonable to conclude that northeast Indian females did not experience any substantial recent loss of genetic diversity, contrary to northeast Indian males.

Similar to the Y-chromosome evidence, mtDNA evidence clearly indicates close genetic affinities between northeast Indian and east Asian groups (Cordaux et al. 2003) with hardly any contribution from nonnortheast Indians to the highly homogeneous northeast Indian mtDNA gene pool. In sum, both Y-chromosome and mtDNA evidence indicate that northeast Indian groups have remained genetically isolated for centuries, without admixing with their close Indian neighbors. The high incidence of genetic traits such as color blindness (e.g., which occurs at more than 10% in Apatani [Jaswal 1975; Singh 1998]) suggests considerable levels of inbreeding in these groups and provides additional evidence for their genetic isolation.

Was Northeast India a Barrier or Corridor for Human Migrations?

The findings of the present study are surprisingly contradictory with the initial prediction of high genetic heterogeneity and diversity in northeast India, based on the fact that the northeast Indian passageway is an important linguistic contact zone, and it is generally believed to have been a key area for migrations between India and East/Southeast Asia (see *Introduction* [Basu et al. 2003]). Genetic analyses of other postulated corridors have shown evidence of clinal variation and/or admixture between populations located adjacent to corridors such as the Nile River Valley (Klings et al. 1999), central Asia (Karafet et al. 2001; Wells et al. 2001; Zerjal et al. 2002; Comas et al. 2004), and eastern Indonesia (Kayser et al. 2000a).

However, there is a strikingly high genetic homogeneity in northeast India, coupled with a remarkable discontinuity in both Y-chromosome and mtDNA variation between northeast India and the rest of India. In the light of the present genetic evidence, we suggest that the northeast Indian passageway acted as a geographic barrier between the Indian subcontinent and East/Southeast Asia, rather than as an important corridor connecting these two major subcontinental areas, at least since the arrival of Tibeto-Burman speakers in northeast India.

Several recent studies have emphasized that the mtDNA and Y-chromosome gene pools of the Indian subcontinent and East/Southeast Asia are related but overall fairly distinct (Kivisild et al. 1999, 2003; Bamshad et al. 2001; Forster et al. 2001; Roychoudhury et al. 2001; Cordaux et al. 2003, 2004). The relatedness of the two gene pools may be interpreted as an indication that the northeast Indian passageway acted as a corridor between the Indian subcontinent and East/Southeast Asia during the initial settlement of the latter area. However, the distinctiveness of the same gene pools argues against considerable gene flow between the two areas for a long period of time, perhaps as long as 30,000 years, as suggested by mtDNA evidence

(Forster et al. 2001). Taken together with the evidence presented in this study, we suggest that the northeast Indian passageway mainly acted as a barrier to migrations during most of modern human history.

Acknowledgments

We are grateful to the original donors of samples. We thank Peter Underhill for kindly providing DNA controls and Isabelle Dupanloup for help with using the ADMIX software. We also thank Silke Brauer for fruitful discussions and Mark Batzer, David Hughes, and Brigitte Pakendorf for critical reading of the manuscript. R.C., G.W., and M.S. were supported by funds from the Max Planck Society, Germany.

Literature Cited

- Abbi, A. 1991. Reduplication in south Asian languages: an areal, typological and historical study. Allied Publishers, New Delhi.
- Bamshad, M., T. Kivisild, W. S. Watkins et al. (18 co-authors). 2001. Genetic evidence on the origins of Indian caste populations. *Genome Res.* **11**:994–1004.
- Bandelt, H. J., P. Forster, and A. Röhl. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**:37–48.
- Basu, A., N. Mukherjee, S. Roy et al. (12 co-authors). 2003. Ethnic India: a genomic view, with special reference to peopling and structure. *Genome Res.* **13**:2277–2290.
- Cann, R. L. 2001. Genetic clues to dispersal in human populations: retracing the past from the present. *Science* **291**:1742–1748.
- Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza. 1994. History and geography of human genes. Princeton University Press, Princeton, NJ.
- Clark, V. J., S. Sivendren, N. Saha, G. R. Bentley, R. Aunger, S. M. Sirajuddin, and M. Stoneking. 2000. The 9-bp deletion between the mitochondrial lysine tRNA and COII genes in tribal populations of India. *Hum. Biol.* **72**:273–285.
- Comas, D., S. Plaza, R. S. Wells, N. Yuldaseva, O. Lao, F. Calafell, and J. Bertranpetit. 2004. Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur. J. Hum. Genet.* **12**:495–504.
- Cordaux, R., R. Aunger, G. Bentley, I. Nasidze, S. M. Sirajuddin, and M. Stoneking. 2004. Independent origins of Indian caste and tribal paternal lineages. *Curr. Biol.* **14**:231–235.
- Cordaux, R., N. Saha, G. R. Bentley, R. Aunger, S. M. Sirajuddin, and M. Stoneking. 2003. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur. J. Hum. Genet.* **11**:253–264.
- Dupanloup, I., and G. Bertorelle. 2001. Inferring admixture proportions from molecular data: extension to any number of parental populations. *Mol. Biol. Evol.* **18**:672–675.
- Etler, D. A. 1996. The fossil evidence for human evolution in Asia. *Annu. Rev. Anthropol.* **25**:275–301.
- Excoffier, L., and S. Schneider. 1999. Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions? *Proc. Natl. Acad. Sci. USA* **96**:10597–10602.
- Forster, P., A. Torroni, C. Renfrew, and A. Röhl. 2001. Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Mol. Biol. Evol.* **18**:1864–1881.
- Hammer, M. F., and S. Horai. 1995. Y chromosomal DNA variation and the peopling of Japan. *Am. J. Hum. Genet.* **56**:951–962.
- Jaswal, I. J. S. 1975. Anomalous color vision among three tribes of Arunachal Pradesh, India. *Anthropologist* **22**:60–64.
- Jobling, M. A., and C. Tyler-Smith. 2003. The human Y-chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet.* **4**:598–612.
- Karafet, T., L. Xu, R. Du, W. Wang, S. Feng, R. S. Wells, A. J. Redd, S. L. Zegura, and M. F. Hammer. 2001. Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am. J. Hum. Genet.* **69**:615–628.
- Kayser, M., S. Brauer, G. Weiss, W. Schiefenhövel, P. A. Underhill, P. Shen, P. Oefner, M. Tomaseo-Ponzetta, and M. Stoneking. 2003. Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am. J. Hum. Genet.* **72**:281–302.
- Kayser, M., S. Brauer, G. Weiss, W. Schiefenhövel, P. A. Underhill, and M. Stoneking. 2001. Independent histories of human Y chromosomes from Melanesia and Australia. *Am. J. Hum. Genet.* **68**:173–190.
- Kayser, M., S. Brauer, G. Weiss, P. A. Underhill, L. Roewer, W. Schiefenhövel, and M. Stoneking. 2000a. Melanesian origin of Polynesian Y chromosomes. *Curr. Biol.* **10**:1237–1246.
- Kayser, M., A. Caglia, D. Corach et al. (30 co-authors). 1997. Evaluation of Y-chromosomal STRs: a multicenter study. *Int. J. Legal Med.* **110**:125–133.
- Kayser, M., L. Roewer, M. Hedman et al. (14 co-authors). 2000b. Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome revealed by direct observation in father/son pairs. *Am. J. Hum. Genet.* **66**:1580–1588.
- Ke, Y., B. Su, X. Song et al. (23 co-authors). 2001. African origin of modern humans in east Asia: a tale of 12,000 Y chromosomes. *Science* **292**:1151–1152.
- Kivisild, T., M. J. Bamshad, K. Kaldma et al. (14 co-authors). 1999. Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr. Biol.* **9**:1331–1334.
- Kivisild, T., S. Rootsi, M. Metspalu et al. (18 co-authors). 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* **72**:313–332.
- Krings, M., A. H. Salem, K. Bauer et al. (13 co-authors). 1999. MtDNA analysis of Nile River Valley populations: a genetic corridor or barrier for migration? *Am. J. Hum. Genet.* **64**:1166–1176.
- Lahr, M. M., and R. Foley. 1994. Multiple dispersals and modern human origins. *Evol. Anthropol.* **3**:48–60.
- Masica, C. P. 1991. The Indo-European languages. Cambridge University Press, Cambridge, UK.
- Misra, V. N. 2001. Prehistoric human colonization of India. *J. Biosci.* **26**:491–531.
- Nei, M., and A. K. Roychoudhury. 1993. Evolutionary relationships of human populations on a global scale. *Mol. Biol. Evol.* **10**:927–943.
- Oota, H., W. Settheetham-Ishida, D. Tiwawech, T. Ishida, and M. Stoneking. 2001. Human mtDNA and Y-chromosome variation is correlated with matrilineal vs. patrilineal residence. *Nat. Genet.* **29**:20–21.
- Quintana-Murci, L., C. Krausz, T. Zerjal et al. (13 co-authors). 2001. Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am. J. Hum. Genet.* **68**:537–542.
- Ramana, G. V., B. Su, L. Jin, L. Singh, N. Wang, P. Underhill, and R. Chakraborty. 2001. 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.* **9**:695–700.
- Roychoudhury, S., S. Roy, A. Basu, R. Banerjee, H. Vishwanathan, M. V. Usha Rani, S. K. Sil, M. Mitra, and P. P.

- Majumder. 2001. Genomic structures and population histories of linguistically distinct tribal groups of India. *Hum. Genet.* **109**:339–350.
- Ruhlen, M. 1991. *A guide to the world's languages*, Vol 1. Stanford University Press, Stanford, Calif.
- Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin ver. 2.000: a software for population genetics data analysis. Genetics and Biochemistry laboratory, University of Geneva, Switzerland.
- Seielstad, M. T., E. Minch, and L. L. Cavalli-Sforza. 1998. Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* **20**:278–280.
- Singh, K. S. 1998. *People of India: India's communities*, Vol IV-VI. Oxford University Press, Delhi, India.
- Su, B., L. Jin, P. Underhill et al. (11 co-authors). 2000*a*. Polynesian origins: insights from the Y chromosome. *Proc. Natl. Acad. Sci. USA* **97**:8225–8228.
- Su, B., C. Xiao, R. Deka et al. (11 co-authors). 2000*b*. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum. Genet.* **107**:582–590.
- Underhill, P. A., G. Passarino, A. A. Lin, P. Shen, M. M. Lahr, R. A. Foley, P. J. Oefner, and L. L. Cavalli-Sforza. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* **65**:43–62.
- Ward, R. H., B. L. Frazier, K. Dew-Jager and S. Pääbo. 1991. Extensive mitochondrial diversity within a single Amerindian tribe. *Proc. Natl. Acad. Sci. USA* **88**:8720–8724.
- Wells, R. S., N. Yuldasheva, R. Ruzibakiev et al. (27 co-authors). 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl. Acad. Sci. USA* **98**:10244–10249.
- Wilson, I., M. Weale, and D. Balding. 2003. Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *J. Roy. Stat. Soc. A* **166**:155–188.
- Wu, X. Z., and F. E. Poirier. 1995. *Human evolution in China*. Oxford University Press, Oxford, UK.
- Zerjal, T., R. S. Wells, N. Yuldasheva, R. Ruzibakiev, and C. Tyler-Smith. 2002. A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. *Am. J. Hum. Genet.* **71**:466–482.

Lisa Matisoo-Smith, Associate Editor

Accepted April 14, 2004