Am. J. Hum. Genet. 56:979–992, 1995

Demographic History of India and mtDNA-Sequence Diversity

Joanna L. Mountain,[1] Joan M. Hebert,[1] Silanjan Bhattacharyya,[2] Peter A. Underhill,[1] Chris Ottolenghi,[3] Madhav Gadgil,[2] L. Luca Cavalli-Sforza[1]

[1]Department of Genetics, Stanford University, Stanford; [2]Centre for Ecological Sciences, Indian Institute of Science, Bangalore; and [3]Villa Saint Michel, Paris

Summary

The demographic history of India was examined by comparing mtDNA sequences obtained from members of three culturally divergent Indian subpopulations (endogamous caste groups). While an inferred tree revealed some clustering according to caste affiliation, there was no clear separation into three genetically distinct groups along caste lines. Comparison of pairwise nucleotide difference distributions, however, did indicate a difference in growth patterns between two of the castes. The Brahmin population appears to have undergone either a rapid expansion or steady growth. The low-ranking Mukri caste, however, may have either maintained a roughly constant population size or undergone multiple bottlenecks during that period. Comparison of the Indian sequences to those obtained from other populations, using a tree, revealed that the Indian sequences, along with all other non-African samples, form a starlike cluster. This cluster may represent a major expansion, possibly originating in southern Asia, taking place at some point after modern humans initially left Africa.

Introduction

The current genetic diversity within a given region is the consequence of that region's demographic history, including the effects of geographic location (with respect to other populated areas), population-size fluctuations, and social structure. India, with a central geographic location, an extended period of human occupation, and a hierarchical social structure, provides unique opportunities for exploring the relationship between genetic diversity and demographic history. Here we describe the diversity among mtDNA sequences of Indian individuals from three groups living in coastal southwestern India. Tree analyses and distributions of pairwise nucleotide differences enabled us to infer demographic histories of these groups. In addition, by comparing these sequences to those of other world populations we inferred the demographic history of these populations within the broader context of human evolutionary history.

Indian society has traditionally been divided into castes, tribal groups, and religious communities. Within the hierarchical Hindu caste system, each caste belongs to one of five *varnas* (in order of decreasing status: Brahmin, Kshatriya, Vaishya, Sudra, and Pancham) (Malhotra 1984). An essentially endogamous *varna* consists of several endogamous caste clusters, which, in turn, include several endogamous castes (Malhotra 1984). Thousands of such castes exist today. Hundreds of endogamous tribal groups also exist. Members of tribal groups of continental India (excluding the northeast Mongoloid groups) tend to be of short stature, with curly hair and dark pigmentation relative to other Indians.

Previous studies of such classical genetic markers as blood groups indicated that Brahmin groups do not associate with one another in inferred trees. Nor do the Brahmin groups consistently associate, in inferred trees, with other caste groups of their region (Karve and Malhotra 1968). Previous studies also revealed that tribal populations are genetically more similar to each other than to other nontribal Indians, but more similar to the nontribal Indians than to morphologically similar populations in Africa and Australia (Roychoudhury 1984). Geographic maps of single alleles for India are unique in that the high degree of microgeographic variation makes the fit of allele-frequency surfaces unusually difficult (Piazza et al. 1981). Cavalli-Sforza et al. (1994) suggest that the widespread net of tribal groups, along with the caste system, creates a multitude of endogamous pockets, leading to this unique pattern.

The three Indian groups examined in this study were (1) the Havik, members of a Brahmin caste from villages within the talukas (administrative regions) of Sirsi, Siddapur, Yellapur, Honnavar, and Kumta, in northwest Karnataka; (2) the Mukri, members of a scheduled caste from several villages within Kumta taluka in northwest Karnataka; and (3) the Kadar, members of a tribal group from the Anapandam settlement of Kadar at Chalkudi range in the Trissur District of Kerala (fig. 1). One additional Havik individual was from Kasargod in the state of Kerala.

Received September 21, 1994; accepted for publication December 30, 1994.

Current address and address for correspondence and reprints: Dr. Joanna L. Mountain, Department of Integrative Biology, University of California, Berkeley, VLSB Room 4151, Berkeley, CA 94720-3140. E-mail: joanna@mws4.biol.berkeley.edu

0002-9297/95/5604-0020$02.00

Figure 1 Map of southwestern India, indicating locations where samples from three endogamous castes were collected. Numbered locations are talukas (administrative units generally covering ~1,000 km^2 and including 20–40 villages): 1, Yellapur; 2, Sirsi; 3, Kumta; 4, Siddapur; 5, Honnavar; and 6, Anapandam (Trissur). Traditionally, villages include families belonging to several different caste groups.

Karnataka is among a small group of Indian states/territories with a large number of culturally diverse communities, possibly harboring caste populations with considerable caste-based occupational specialization and more rigid cultural practices than in other regions of India (Joshi et al. 1993). The coastal and hill region of the state of Karnataka considered here was among the last areas to be brought under settled cultivation. While agriculture and animal husbandry reached India possibly as early as 8,000 years before the present, these practices reached coastal Karnataka only ~3,000 years ago.

The Havik are among the Brahmin endogamous caste groups, which have the highest status in the caste system. They are thought to have immigrated ~1,300 years ago to the 20,000 km^2 area where they are now concentrated. Today the group includes ~100,000 individuals who are predominantly specialized cultivators of betel nuts and spices; priests; and white-collar workers. The Havik depend very little on hunting and gathering and are the only endogamous caste group in the area with access to advanced horticultural techniques.

The Mukri are members of a scheduled caste. Individuals belonging to such castes, ranked at the lowest level in the caste hierarchy, were previously known as "untouchables." The Mukri consist of 8,000–10,000 individuals who survive through hunting, gathering, fishing, or other unskilled labor in a 2,000 km^2 area. They have provided unskilled labor in the agrarian society for the past 2,500 years and have only recently begun to practice agriculture. The Mukri depend on hunting and gathering to a greater extent than any other endogamous group in the region and may have done so for a long time; they have no oral tradition of migration and were probably the local hunter-gatherers when the Havik immigrated to this area (Gadgil and Iyer 1989; Gadgil and Guha 1993).

The Kadar are a group of hunter-gatherers living in a mountainous region of the district of Kerala. They are members of a tribal group, separate from the Hindu caste system. Elsewhere when the Kadar were compared to other Indian populations through tree inference from classical genetic markers, they appeared as the extreme outlier (Cavalli-Sforza et al. 1994). In the present study, very few individuals (seven) were typed; while their sequences are provided, few conclusions regarding the Kadar population can be drawn.

Two segments of the highly variable "D"-loop region of the mtDNA genome were sequenced. Human mtDNA is maternally inherited (Giles et al. 1980), although there are indications of a minor paternal contribution in other organisms (Gyllensten et al. 1991; Hoeh et al. 1991). This 16.5-kb circular molecule is present in much higher copy number (Robin and Wong 1988) and mutates at a higher rate (Wilson et al. 1985) than nuclear DNA. The rate of nucleotide substitution per site per year for nuclear DNA is estimated to average 4.7×10^{-9} (Nei 1987, p. 147). This corresponds to a rate of $\sim 10^{-7}$ substitutions per site per generation, assuming a generation length of 20 years for humans. The substitution rate for mtDNA is estimated to be ~10 times higher than for nuclear DNA, on the order of 10^{-6} (Brown et al. 1979). The D-loop region contains sites that regulate mtDNA replication and transcription. Two portions of this region are particularly variable and have therefore been denoted the "hypervariable" regions. The mutation rate in these regions appears to be highly variable across sites (Aquadro and Greenberg 1983; Wakeley 1993). The average rate may be an order of magnitude higher than for other regions of the mitochondrial genome.

We have considered only individuals for which these two hypervariable segments of the control region have been sequenced, thereby maximizing the number of sites for all comparisons. While similar analyses have been carried out on more individuals sequenced for only one segment, consideration of both segments should lead to more robust results in both tree inference and pairwise-difference distribution comparisons.

Subjects, Material, and Methods

Sample Collection

The locations of talukas (generally including 20–40 villages) where samples were collected are indicated on the map (fig. 1). Because the Mukri reside primarily in Kumta taluka, all Mukri samples were obtained from this region. The Havik reside in a broader geographic

region, and samples were obtained from an area including five talukas. All individuals sampled are believed to be maternally unrelated for ≥2 previous generations. Field sampling was performed as follows: sterile, flat toothpicks (usually 6) were used to scrape the buccal cheek area and tongue. Rigorous controls were performed to ensure that the sampling materials were free of any contamination. The scraped cells were inserted into a sterile tube containing 0.6 ml "PCR buffer + nonionic detergents" (Higuchi 1989) + 1% (w/v) sodium azide.

Scalp hairs were plucked using a new pair of rubber gloves for each person sampled. They were stored dry in sterile Eppendorf tubes. In the lab the hairs were rinsed with sterile, distilled water and stored in 100% ethanol. The portion of the hair that was touched during sampling was removed by burning.

mtDNA from each of the six individuals who were involved in the sample collection or PCR amplification was sequenced as a control against undetected contamination. Each of these sequences was determined to be different from any of the Havik, Mukri, and Kadar sequences.

Amplification and Purification

To extract DNA, buccal cells were digested with proteinase K (Boehringer-Mannheim) as follows: 3 μl of a 10 mg/ml (aqueous) solution of proteinase K was added per 100 μl of sample. The sample was heated at 55°C for 1 h, and then the proteinase K was inactivated by heat treatment for 10 min at 95°C. Either 4 μl or 20 μl of extract was used in a 100-μl PCR reaction. The root portions of 1–3 hairs of a single individual were air dried and snipped into 0.5 ml "PCR Buffer + non-ionic detergent." Proteinase K digestion was performed as described. Fifty microliters of the hair root extract was used in a 100-μl PCR reaction.

Amplifications were carried out using a Perkin-Elmer-Cetus DNA Thermal Cycler. Each reaction contained 20 mM Tris pH 8.3, 1.5 mM $MgCl_2$, 25 mM KCl, 0.05% Tween-20, 0.2 μM each primer, 50 μM each dNTP, 100 μg/ml nuclease-free acetylated BSA (Promega or New England BioLabs), and 2.5 U *Taq* polymerase (Perkin-Elmer-Cetus). Segment I (404 bp excluding primers) was amplified with L15996 (5′-CTCCACCATTAGCACCCAAAGC-3′) and H16401 (5′-TGATTTCACGGAGGATGGTG-3′). Segment II (378 bp excluding primers) was amplified with L29 (5′-GGTCTATCACCCTATTAACCAC-3′) and H408 (5′-CTGTTAAAAGTGCATACCGCCA-3′). The numbers indicate the 3′ base according to the reference sequence (Anderson et al. 1981), and the L and H refer to "light" and "heavy" strands, respectively. A negative control was included with each amplification.

The thermal regimen consisted of 30 cycles of denaturation at 94°C for 45 s, 56°C for 1 min, and 74°C for 1 min, followed by a final 7 min extension at 72°C. This generally resulted in a single product for segment II but often resulted in multiple fragments for segment I. Consequently, segment I products had to be purified on a low-melting-point agarose gel as described below. Subsequently, amplification conditions were optimized, resulting in a single specific segment I product. This was achieved by increasing the annealing temperature to 66°C and using a "hot start" method (D'Aquila et al. 1991) in which all reagents except for *Taq* polymerase were preheated prior to the addition of polymerase.

There appeared to be an inhibitor of *Taq* polymerase present in some of the buccal mucosa samples. The use of a 4 μl instead of a 20 μl sample often led to successful amplification. In other cases, however, it was necessary to extract the total DNA from these samples before any amplification could be achieved. This was performed as follows: the sample was extracted with an equivalent volume of equilibrated phenol. The aqueous layer was removed, and the phenol was back-extracted with TE buffer (10 mM Tris pH 8.0 and 1 mM EDTA). The two aqueous layers were pooled and extracted once with phenol:chloroform:iso-amyl alcohol (50:48:2). The aqueous layer from this was removed and extracted once with chloroform:iso-amyl alcohol (24:1). Finally, the DNA was ethanol-precipitated from the aqueous layer and was resuspended in 20 μl TE. A negative control was carried through this procedure each time it was performed.

After amplification 10 μl of each product was run on a minigel (2% agarose) to determine amplification success, by using a 123-bp ladder (Gibco-BRL) as a marker. When a single band of appropriate size was detected, the product was purified using either a Pharmacia Miniprep Spun Column (Sephacryl S-400) or a Promega Magic PCR prep kit. When the amplification contained secondary bands the reaction was loaded onto a 2% low-melting-point agarose gel. The desired band was excised and melted, and the DNA was then purified using a Promega Magic PCR prep kit.

Sequencing

Double-stranded sequencing of the purified PCR product was performed using either conventional methods or fluorescent dye labeling and an Applied Biosystems 373A DNA sequencer. For conventional sequencing, Sequenase Version 2.0 (U.S. Biochemical) was used with 2-deoxyadenosine 5′-(a-(35^S) thio) triphosphate (Amersham). Denaturation of the PCR product was achieved by boiling the template plus sequencing primer in annealing buffer, followed by immersion in a dry-ice/ethanol bath. Products were electrophoresed through 6% polyacrylamide, 7 M urea on a Biorad Sequigen Apparatus for 2.5 or 5 h; then the gel was dried and exposed to Kodak XAR-5 film. For ABI 373A analyzed sequences, DyeDeoxy terminator cycle sequencing was

done using a Perkin-Elmer-Cetus 9600 GeneAmp PCR System, following recommended sequencing kit protocols.

In general, segment II was sequenced with conventional methods, although many of these sequences were repeated on an ABI 373A, and segment I was sequenced using the automated sequencer. Each segment was sequenced in both directions by using the amplification primers as sequencing primers. For segment I, it was sometimes necessary to do an additional sequencing reaction with a third, internal primer, L16209 (5′-CCCCATGCTTACAAGCAAGT-3′). When base position 16189 was a C rather than a T, the sequence quality beyond this region was ambiguous on both strands. Use of L16209, which anneals just beyond this region, made determination of a complete composite sequence possible.

Sequence Analyses

Sequences were aligned using CLUSTAL V (Higgins and Sharp 1988) and by hand. Distances between sequences were estimated with DNADIST of the PHYLIP software package (Felsenstein 1989) using Kimura's two-parameter model, with transitions assumed to occur at 10 times the rate of transversions (Kimura 1980). Phylogenetic analyses were carried out using NEIGHBOR; in some cases, in particular for trees relating hundreds of sequences, branch lengths were inferred to be of negative length. These were all relatively small in magnitude, however, and were therefore represented by branches of length zero. Trees were compared using Templeton's Delta Q test, in the PHYLIP software package. Trees were drawn using the DRAWTREE and DRAWGRAM programs; bootstrap data sets were obtained using SEQBOOT; and consensus trees were obtained using the CONSENSE program provided with the PHYLIP software package (Felsenstein 1989).

The time and extent of ancient demographic expansions can be inferred from distributions of pairwise nucleotide differences, or mismatch distributions, if the distributions are smooth and nearly Poisson (Slatkin and Hudson 1991; Rogers and Harpending 1992; Rogers, in press). Slatkin and Hudson consider an exponentially growing population in which the product of the current population size and the growth rate is substantially larger than 1. They assume that exponential growth from population size N_1 to size N_0 has taken place at a rate, r, over a number of generations, t. In this case the distribution of pairwise differences is expected to be approximately Poisson. They demonstrate that one can estimate r, given the mean number of pairwise differences between sequences, and assuming that N_0 and m, the mutation rate per site per generation, are known. If it is assumed that m is known, one can also estimate t: $t = i/(2m)$, where i is the mean number of pairwise differences per site.

Rogers and Harpending (1992) use either two or three parameters to characterize the expansion of a population, given a pairwise-difference distribution. They consider a model of sudden expansion: at $t = 0$ a population grows suddenly from effective female population size N_0 to female population size N_1 and is observed t generations later. Rogers (in press) gives a method for estimating $\theta_0 = 2\mu N_0$ and $\tau = 2\mu t$, from two moments, and $\theta_0 = 2\mu N_0$, $\theta_1 = 2\mu N_1$ and $\tau = 2\mu t$, from three moments of the pairwise-difference distribution (μ is the mutation rate). We considered mutation rates of 10^{-5}–10^{-6} per site per generation. While estimates of mutation rates for the mtDNA D-loop are very rough, this range may not be unrealistic (Nei 1987, p. 147). Demographic expansion parameters were estimated from pairwise-difference distributions using either MATHEMATICA (Wolfram 1988) or a computer program provided by Alan Rogers. Estimates of the unevenness, or "raggedness," of mismatch distributions were obtained using the approach of Harpending et al. (1993).

Results

Sequence Variability among Indians

Sequences for both variable segments were obtained for 43 individuals of the Mukri population, 48 individuals of the Havik population, 7 individuals of the Kadar population, and 7 Indian and European controls (fig. 2). Among the 791 sites, 86 were found to be polymorphic within this sample of 98 individuals (excluding controls). Of the variable sites, 62 are in segment I and 24 are in segment II. Transitions alone were observed at 73 sites; the incidence of pyrimidine transitions was roughly twice that of purine transitions (C↔T: 50 vs. A↔G: 23). Transversions alone were observed at six sites (A↔T: 3; C↔G: 1; and A↔C: 2). At one position both transitions and transversions were observed. Insertion/deletion events were observed at six sites.

The proportion of distinct sequences differed among the three groups of Indians. Eighteen distinct sequences were detected among the 43 samples from Mukri Indians, 37 distinct sequences were detected among the 48 Havik samples, and 2 distinct sequences were detected among the 7 Kadar samples. Virtual heterozygosity, or haplotype diversity (Hartl and Clark 1989), was highest for the Havik (.97), less high for the Mukri (.86), and lowest for the Kadar (.41). The value for three groups combined was .96. These values were reflected in the inferred trees, discussed below.

Tree Inference

In order to further understand the history of these Indian lineages, we inferred trees according to the neighbor-joining algorithm (Saitou and Nei 1987). The tree relating sequences of 106 individuals (Havik, Kadar, Mukri, controls, plus reference) is given in figure 3. A

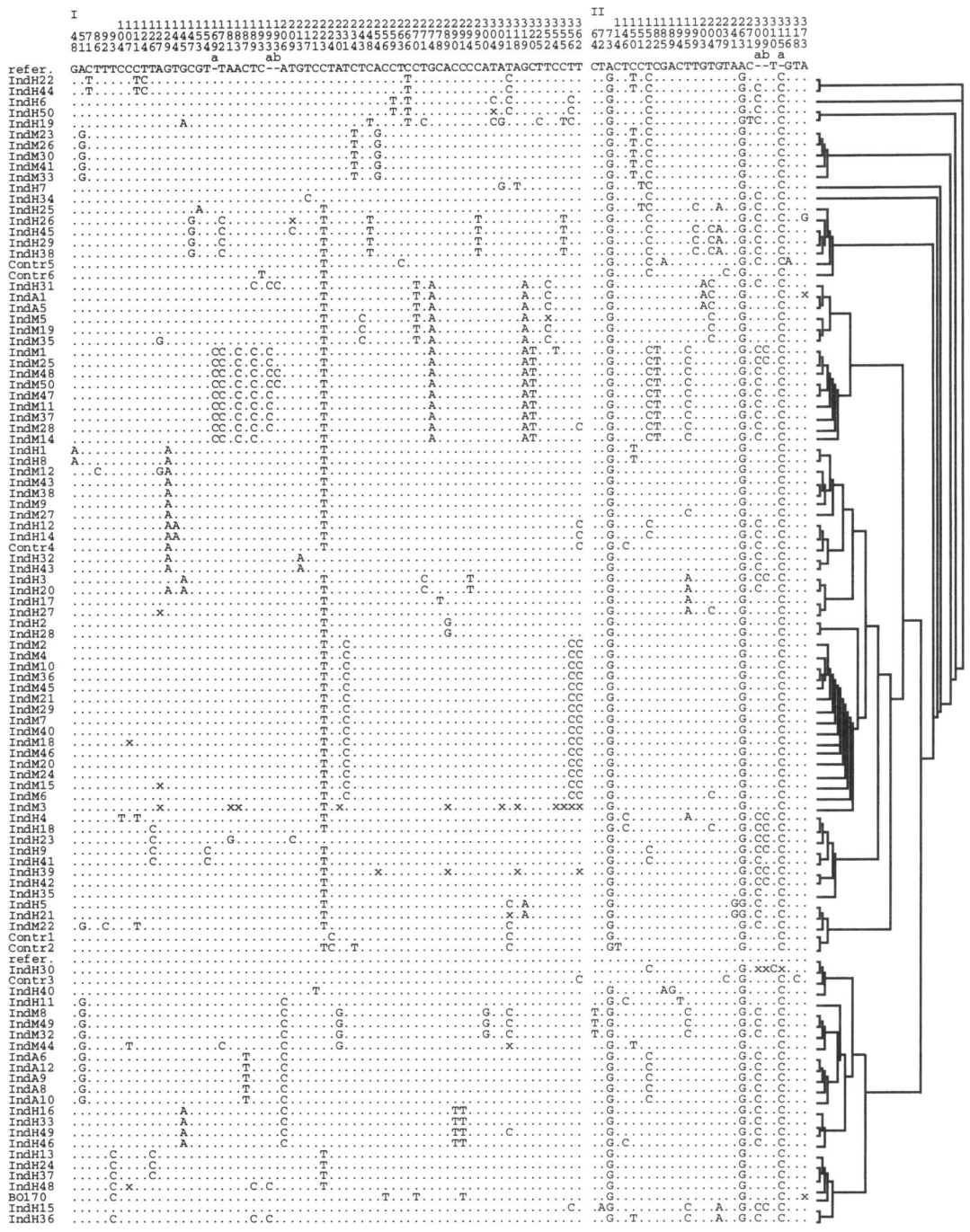

Figure 2 Variable positions for 105 mitochondrial control region sequences (43 Mukri [IndM], 48 Havik [IndH], 7 Kadar [IndA], plus 7 Indian and European controls) and neighbor-joining tree inferred from these sequences. Sequences were aligned (CLUSTAL V, and by hand) with a reference sequence (Anderson et al. 1981). A total of 791 nt sites were considered for the analyses (15997–16401 for segment I and 29–408 for segment II, plus six possible points of insertion). Only the last three digits are given for segment I. When <791 bases were sequenced or when a base could not be called unambiguously (denoted by x), individuals were compared using only sites with data available for both sequences. Sequences were sorted according to the tree shown in fig. 3, were rooted arbitrarily, and were redrawn here. The tree is topologically that shown in fig. 3; in this version, however, branch lengths do not correspond to genetic distance.

neighbor-joining tree inferred excluding control sequences is not significantly longer than the maximum-parsimony tree inferred from the same data set (tree not shown) according to Templeton's Delta Q test: the neighbor-joining tree requires 161 steps, while the maximum-parsimony tree requires 157 steps. Sequences from all three of the endogamous groups are found scattered throughout the neighbor-joining tree and maximum-

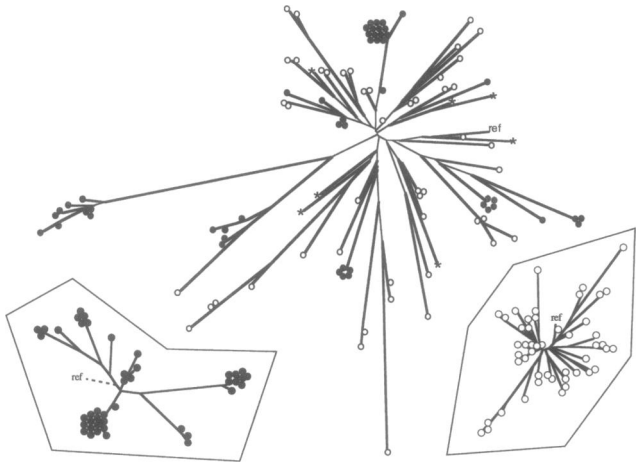

Figure 3 Tree inferred from sequence data of fig. 2. Blackened circles (●) represent Mukri; Patterned circles (❁) represent Kadar; Unblackened circles (○) represent Havik; and asterisks (*) represent controls. Genetic distances among the 106 sequences (including reference sequence) were obtained using Gendist of the PHYlogeny Inference Package (PHYLIP; Felsenstein 1989). Kimura's two-parameter model (Kimura 1980) was assumed, and the ratio of transitions to transversions was assumed to be 10:1. (The resulting trees did not appear to vary greatly when this ratio was varied from 2 to 10; data not shown.) Trees were inferred from these genetic distances according to the neighbor-joining algorithm (Saitou and Nei 1987) using PHYLIP's Neighbor routine. Some branch lengths were inferred to be negative, indicating a lack of fit of the tree to the data. Absolute values, however, were small except for one case, and such branches were represented as having length zero (Saitou and Nei 1987). The sequence of IndM3 led to a large negative branch, possibly because bases of many sites could not be determined unambiguously for this sequence. Robustness of trees was examined by bootstrapping (Efron 1982; Felsenstein 1985) the 791 sites 100 times, inferring neighbor-joining trees for each bootstrap sample, and comparing their consensus with the original neighbor-joining tree. A consensus tree was obtained using the CONSENSE program of PHYLIP (Felsenstein 1989). Branches found in >50% of 100 neighbor-joining bootstraps are indicated by thick lines.

parsimony trees. Such intermixing of lineages suggests that either these populations have undergone extensive mixing or that the mtDNA lineages represented here arose prior to the split among the three populations.

In order to test the significance of the intermixing of lineages from the different populations, we compared the lengths of trees allowing such intermixing with the maximum-parsimony tree constrained so that all individuals of each population cluster together. Both the maximum-parsimony and the neighbor-joining tree are significantly shorter (Templeton's Delta Q test) than the maximum-parsimony tree constrained so that there is no intermixing (185 steps). Thus, the cultural identification of these individuals is inconsistent with the genetic grouping.

To some extent, however, sequences do cluster according to the Indian caste groups; individuals almost always form clusters with other individuals of their own group. A cluster can be defined as a set of sequences that are separated from all other sequences by a branch. All Mukri sequences except for one fall into six clusters. These six clusters include only Mukri sequences with one exception: in one case five Mukri sequences cluster with two Havik sequences.

A bootstrap test of robustness of this tree indicated that, as expected, lengths of longer branches of the tree are significantly different from zero, while lengths of shorter branches, particularly toward the center of the tree, are not. Branches of length significantly greater then zero, according to a consensus of bootstraps, are indicated by heavy lines in figure 3. The lack of support for central branches suggests that this tree is starlike, a pattern that arises when a population has undergone an expansion (Slatkin and Hudson 1991).

Even in the case where Mukri and Havik individuals are found together in a cluster, the Mukri individuals are well separated (by a branch supported by bootstraps) from the Havik. While the Havik sequences also cluster to some extent, robust clusters tend to be small (2–4 sequences). The Kadar fall into two well-separated clusters. While in some cases discordance between gene and population genealogies may reflect admixture, there is little evidence here of high levels of recent mixture between the Havik and other subpopulations: specifically, Mukri and Kadar clusters tend to be well separated from the Havik lineages.

The topologies of the two trees inferred for the Havik and Mukri subpopulations separately (fig. 3; see insets) differ substantially, as might be expected from the combined tree. The tree inferred from the Havik sequences is more starlike: many lineages branch at roughly the same point in the center of the tree. The tree inferred from the Mukri sequences reveals most sequences attaching to a few central branches. Bootstrap analyses of these trees indicate that the central branches of the Mukri tree are much more reliable than those of the Havik tree. The demographic history behind these differing topologies is explored below, through distributions of pairwise differences.

Pairwise-Difference Distributions

Much of the tree shown in figure 3 is robust according to the test described above, giving some idea of the genealogy of these Indian mtDNA lineages. The demographic history of the Indian groups represented here can be examined further by studying the distribution of nucleotide differences among these sequences considered two at a time (Slatkin and Hudson 1991; Rogers and Harpending 1992; Rogers, in press). These distributions are expected to be multimodal for populations that have retained a constant population size over some time. A smoother, unimodal distribution is expected for a population that has undergone either a sudden expansion or exponential growth over time (Slatkin and Hudson 1991; Rogers and Harpending 1992).

The distribution of pairwise differences for the Indian

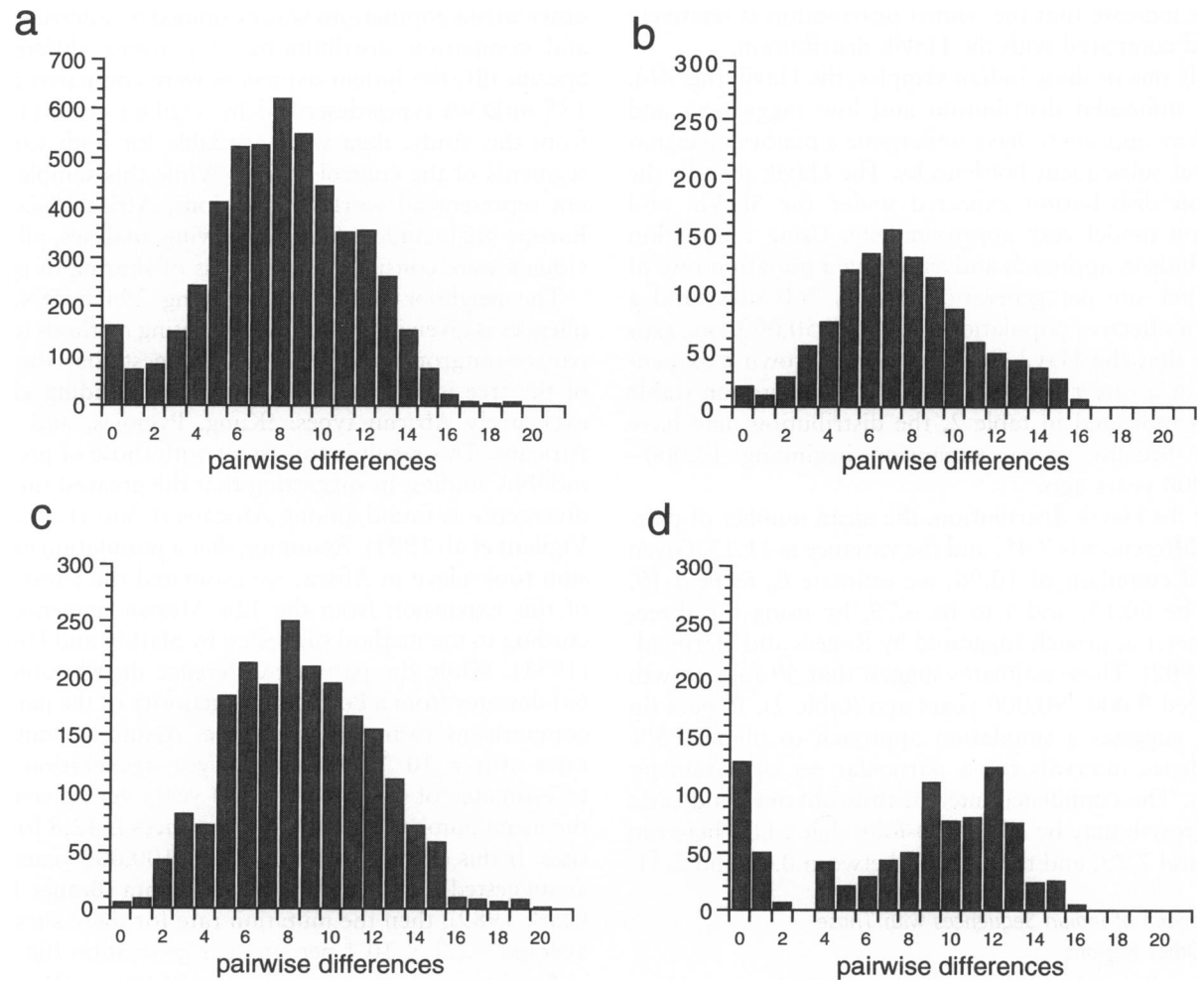

Figure 4 Pairwise-difference distributions for the following: *a*, 101 Indian sequences—Mukri, Havik, Kadar, plus three controls (mean = 8.02 over 716 sites); *b*, 48 Havik samples (mean = 7.49 over 750 sites); *c*, 48 Havik samples compared to 43 Mukri samples (mean = 8.61 over 749 sites); and *d*, 43 Mukri samples (mean = 7.93 over 748 sites). Vertical axis is number of pairs separated by a given number of differences. Insertion/deletion events were not considered to be differences.

samples (fig. 4*a*) clearly deviates from a Poisson, with a small mode at zero. The majority of the pairwise comparisons, however, do form a fairly smooth distribution. In order to further understand this distribution we examined the distributions for the Havik and Mukri separately (fig. 4*b* and *d*) as well as the distribution of differences between the two subpopulations (fig. 4*c*). These distributions suggest a contrast between the demographic histories of the two subpopulations. Pairwise differences among the Mukri samples have a multimodal distribution, the signature of a roughly constant population size, while the Havik sequences have a single mode, suggesting population growth. The Kadar sample is too small to provide a clear indication of a distribution. Harpending et al. (1993) described a statistic, "raggedness," which measures the unevenness of a pairwise-difference, or mismatch, distribution Simulations suggest that raggedness values of .0004 ± .0006 reflect expansions while values of .40 ± .35 imply "raggedness" and correspond to a roughly constant population size over time. Values for the Havik and Mukri are given in table 1. While the behavior of this statistic has been explored through simulation, a theoretical foundation for its application has not yet been described, and standard errors are unavailable. Nonetheless, the observed

Table 1

Values for the Statistical Measure of Raggedness Suggested by Harpending et al. (1993)

Population	Segment I	Segment II	Segments I + II
Havik	.0098	.0292	.0057
Mukri	.0750	.0825	.0242

NOTE.—Larger values indicate that a distribution is more uneven, or ragged. Segment I includes 404 bp, and segment II includes 378 bp of the mtDNA D-Loop region.

values indicate that the Mukri distribution is relatively ragged compared with the Havik distribution.

Only one of these Indian samples, the Havik (fig. 4*b*), has a unimodal distribution and low raggedness and therefore appears to have undergone a major expansion without subsequent bottlenecks. The Havik data fit the Poisson distribution expected under the Slatkin and Hudson model very approximately. Using the Slatkin and Hudson approach and assuming a mutation rate of 10^{-5} per site per generation (across 760 sites) and a current effective population size N_0 of 50,000, one estimates that the Havik population has grown exponentially at a rate r of ~1% or less per generation (table 2). As indicated in table 2, the distribution may have arisen because of an expansion beginning 10,000–100,000 years ago.

For the Havik distribution, the mean number of pairwise differences is 7.48, and the variance is 11.23. Given a third cumulant of 10.96, we estimate θ_0 to be 1.19, θ_1 to be 60.13, and τ to be 6.79, by using the three-parameter approach suggested by Rogers and Harpending (1992). These estimates suggest that 50-fold growth occurred 9,000–90,000 years ago (table 2). Rogers (in press) suggests a simulation approach to obtain 95% confidence intervals for a particular set of parameter values. The confidence interval thus obtained indicates that growth may be 30–1,000-fold, that τ falls between 4.78 and 7.79, and that θ_0 falls between 0.12 and 2.51.

Comparison of Indian Sequences with Those from Other Regions

The relationship between Indian mtDNA control region sequences and those sequences of individuals from other world populations was examined by inferring trees and comparing distributions of pairwise differences. Specifically, the Indian sequences were compared to the 135 mtDNA types described by Vigilant et al. (1991); from this study, data were available for both variable segments of the control region. While this sample does not represent all world populations, Africa, Asia, and Europe are included. In the following analyses, all individuals were considered regardless of sharing of types.

The neighbor-joining tree relating 294 mtDNA sequences is given in figure 5. Two rooting methods (chimpanzee-outgroup and midpoint) suggest that the root of the tree is at M_1, within a cluster including almost exclusively African types: !Kung, Pygmies, and other Africans. This result is consistent with those of previous mtDNA studies, in suggesting that the greatest mtDNA divergence is found among Africans (Cann et al. 1987; Vigilant et al. 1991). Assuming that a population expansion took place in Africa, we estimated the parameters of this expansion from the 128 African sequences according to the method suggested by Slatkin and Hudson (1991). While the pairwise-difference distribution (fig. 6*a*) deviates from a Poisson, the majority of the pairwise comparisons form a single mode. Assuming mutation rates of $\mu = 10^{-5}$–10^{-6} and 20 years/generation leads to estimates of 20,000–200,000 years ago, given that the mean number of pairwise differences is 12.8 for 660 sites. If this expansion took place ~100,000 years ago, as suggested by some archaeological data (Brauer 1989; Clark 1989), then the mutation rate for these sites may average $\sim 2.0 \times 10^{-6}$ per site, per generation (fig. 6*a*).

As shown in the tree, all non-African mtDNA sequences (except for two Chinese sequences) along with about one-third of all African sequences form a single, large starlike cluster, centered at M_2. This starlike cluster may indicate that a major population expansion took place as or after humans left Africa and began occupying parts of Eurasia. The distribution of pairwise nucleotide differences for the 164 non-African sequences (Indians, Europeans, Asians, New Guineans, and one Australian), comprising most of this starlike cluster, is given in figure 6*b*. Using the approach discussed by Slatkin and Hudson (1991), the time of the putative non-African expansion is estimated to have been between 13,000–130,000 years ago, given the mean number of pairwise differences of 8.96 for ~695 sites, and assuming $\mu = 10^{-5}$–10^{-6} per site, per generation; 20 years per generation. If we assume a mutation rate of 2.0×10^{-6}, estimated above, we estimate that the putative M_2, or non-African, expansion took place ~65,000 years ago.

Table 3 gives estimates of the θ_0 and θ_1 values that correspond to the initial and final population sizes in the Rogers and Harpending model. While the θ_0 values are very similar, the value of θ_1, corresponding to the population size after the expansion, for the non-African distribution, is essentially twice that of the African distribution.

Table 2

Parameter Estimates for Hypothesized Havik Expansion

Method	Parameter[b]	μ[a] 10^{-5}	μ[a] 10^{-6}
Slatkin and Hudson (1991)[c]	T	9,982	99,823
	r	1.16	.00055
Rogers (in press) two moment	T	7,404	74,043
	N_0	129	1289
Rogers (in press) three moment	T	9,059	90,586
	N_0	79	793
	N_1	4,008	40,088

NOTE.—An average of 7.49 differences over 750 sites was observed (see fig. 4). The distribution of pairwise differences for the Havik was characterized according to methods suggested by Slatkin and Hudson (1991) and by Rogers (in press).

[a] μ = mutation rate per site per generation; 1 generation = 20 years.

[b] T = time, in years; r = growth rate per generation (%); N_0 = initial effective female population size; and N_1 = final population size.

[c] N_1 assumed to be 5×10^4.

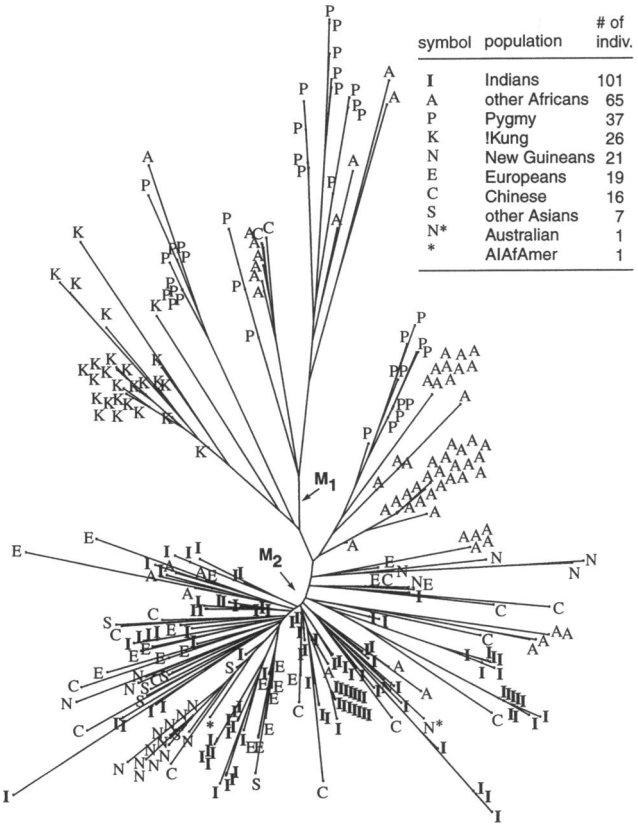

Figure 5 Tree inferred from sequence data of fig. 2, plus sequences described by Vigilant et al. (1991). Sequences were aligned (CLUSTAL V, and by hand) with a reference sequence (Anderson et al. 1981). A total of 745 nt sites were considered for the analyses (15997–16393 for segment I and 68–408 for segment II—numbering according to Anderson et al. [1981]—plus 11 possible points of insertion). Genetic distances among the 294 sequences (101 Indians including 3 controls, 4 European controls, 1 reference sequence, and 188 sequences from Vigilant; repeated mtDNA types were considered individually) were obtained using Gendist of the Phylogeny inference package (PHYLIP; Felsenstein 1989). Kimura's two-parameter model was assumed (Kimura 1980), and the ratio of transitions:transversions was assumed to be 10:1. The tree was inferred from these genetic distances according to the neighbor-joining algorithm (Saitou and Nei 1987) using PHYLIP's Neighbor routine. Some branch lengths were inferred to be negative, indicating a lack of fit of the tree to the data. Absolute values, however, were small, and such branches were represented as having length zero (Saitou and Nei 1987). M_1 indicates the root of the tree as inferred by midpoint and chimpanzee-outgroup rooting methods. M_2 indicates the center of a starlike cluster.

Relationships among world populations can be inferred indirectly from the mean pairwise differences within and between populations. A simple genetic distance among eight populations is given in table 4. From these the tree shown in figure 7 has been inferred. Diversity among African populations is apparent, while subclustering among non-African populations is difficult to detect. Bootstrap proportions given in figure 7 indicate that Indians cluster with Europeans and Chinese, connected by relatively short branches.

Discussion

Methods

While we can infer a tree with some confidence for populations that have maintained a fairly constant pop-

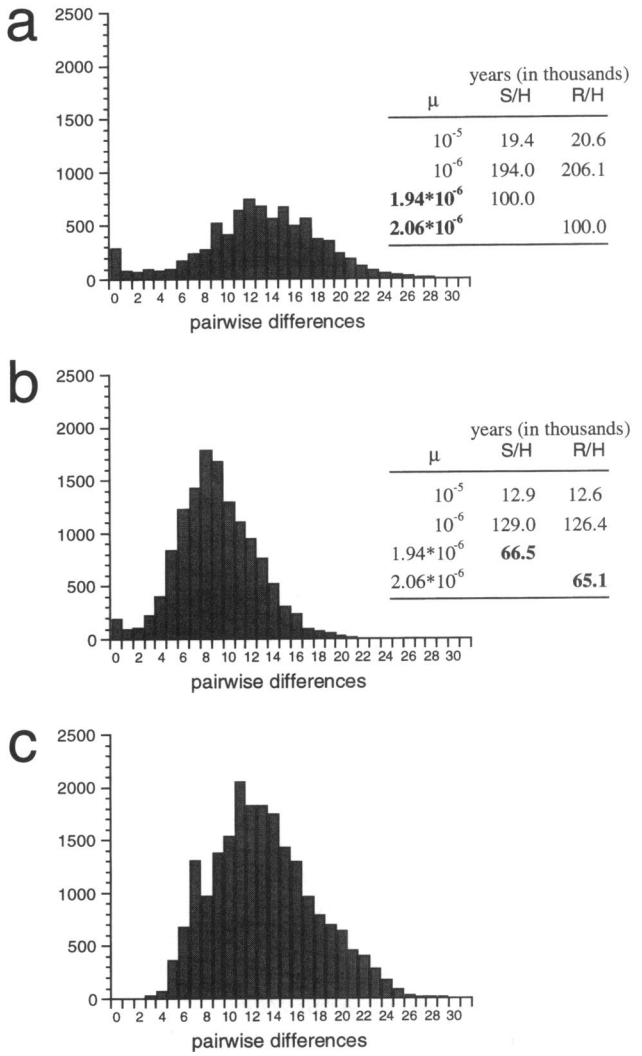

Figure 6 Distributions of number of nucleotide differences between pairs of sequences. μ = mutation rate per site, per generation; years: corresponding estimate of time of expansion. S/H represents estimates of mutation rate or time according to Slatkin and Hudson (1991). R/H represents estimates of mutation rate or time according to Rogers and Harpending (1992). *a,* Distribution for 128 African sequences—with a mean of 12.8 over 660 sites—branching from M_1 in the tree shown in fig. 5. This may correspond to an initial expansion of modern humans from within Africa. Bold values are estimates of mutation rate, assuming that a modern human population began expanding 100,000 years ago. *b,* Distribution for the 164 non-African sequences—with a mean of 8.96 over 695 sites—that form a starlike cluster centered at M_2 in the tree shown in fig. 5. This may correspond to a later expansion of modern humans. Bold values are estimates of time of second major expansion (of non-Africans), assuming mutation rate determined in *a. c,* Between-group distribution: number of nucleotide differences for African/non-African pairs, with a mean of 13.04 over 669 sites.

Table 3

Estimates of Expansion Parameters according to the Method of Rogers and Harpending (1992)

Parameter	Africans	Non-Africans
τ[a]	13.6	8.34
θ_0[b]	1.09	.97
θ_1[c]	53.76	117.88

NOTE.—Estimates of time of expansions are given in fig. 6. No estimates of population sizes are given because these would be overly approximate: parameter estimates are given here for comparison only.

[a] $\tau = 2\ \mu t$, where t is the number of generations since the expansion and μ is the mutation rate.

[b] $\theta_0 = 2\mu N_0$, where N_0 is the initial size of the population.

[c] $\theta_1 = 2\mu N_1$, where N_1 is the size of the population after the putative expansion.

ulation size, a population that has grown either suddenly or steadily over a number of generations leads to a starlike genealogy (Slatkin and Hudson 1991), wherein many lineages branch off at roughly one point in the tree. Such a genealogy may be difficult to reconstruct: the order of the branching (or coalescent) events near that point may be impossible to infer. As in the tree shown in figure 3, central branches may be poorly supported. We are therefore able to infer the tree for the Mukri with more confidence than we can for the Havik. The degree to which a tree is starlike can be seen in the distribution of pairwise differences: a starlike tree corresponds to a narrow, unimodal distribution.

The assumption that expansions occurred suddenly is unrealistic. Rogers and Harpending (1992) indicate, however, that their model provides a good fit to the data even when the actual history is one of continued exponential growth. Rogers (in press) also counters various other criticisms of this approach, including the problem of variable mutation rates across sites. He estimates that such variation introduces an error of ~3%.

Population Histories: Havik versus Mukri

While the overall pattern of pairwise differences for the Indian sequences is somewhat rough (fig. 4*a*), the unimodal distribution indicates that an expansion has occurred. This pattern obscures the difference between the patterns for the Havik and Mukri subpopulations. When the patterns of the Havik and Mukri are examined separately, they reveal two demographic patterns. The Havik appear to have undergone an expansion. The Mukri, on the other hand, appear to have maintained a constant population size, as suggested by the multimodal distribution of pairwise differences. While there are no firm historical data to show that the Havik population has been growing more rapidly in comparison with the Mukri population, there is abundant evidence that over the few thousand years the Brahmin groups had far higher levels of resource access than did untouchable groups.

Other authors, including Vidyarthi (1983) and Cavalli-Sforza et al. (1994), have suggested that there are at least four major components to the genetic structure of India. The first corresponds to the early, Paleolithic occupation of India. The Mukri may be descendants of these earliest inhabitants of southwest India. The second corresponds to the migration of farmers, possibly speaking proto-Dravidian languages, from western Iran during the Neolithic. The third component corresponds to the arrival of Indo-European speakers roughly 3,500 years ago. The ancestors of the Havik may have been part of the second or third components. Archaeological evidence suggests that cultivation spread into India from two directions. The first wave arrived 6,000–7,000 years ago in northwestern India, from its origin in the Near East. The second wave reached eastern India 5,000–6,000 years ago, probably from the independent center of origin of agriculture in China (Herre and Rohrs 1977).

The hierarchical caste society crystallized fairly recently, ~2,000 years ago. The individual caste groups probably arose, however, from endogamous tribelike groups of the earlier society (Karve 1961). Groups may have initially found themselves assigned to the various hierarchical levels based on their roles in society. This role was undoubtedly correlated with the group's con-

Table 4

Simple Distances between Populations (× 10,000)

Population	!Kung	Pygmy	Other African	European	Indian	Chinese	Other Asian
Pygmy	85						
Other African	88	34					
European	112	68	29				
Indian	110	66	28	10			
Chinese	96	53	11	5	5		
Other Asian	118	75	30	7	16	6	
New Guinean	113	81	43	18	28	17	4

NOTE.—$D_b - (D_{w1} + D_{w2})/2$, where D_b = mean number of pairwise differences between sequences from the two different samples, and D_{wi} = mean number of pairwise differences within sample i.

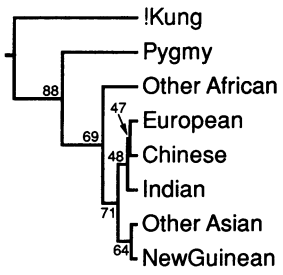

Figure 7 Tree inferred according to average linkage algorithm (unweighted pair-group method using arithmetic averages; see Sokal and Michener [1958]) from simple genetic distances given in table 2. Values at nodes indicate number of 100 bootstrap trees presenting cluster to the right of node. Tree inferred according to neighbor-joining method (Saitou and Nei 1987) (not shown) indicated similar relationships among populations.

trol over technologies, i.e., the degree to which they were able to generate surplus resources and to acquire surplus resources from other groups. At the time of formation of the caste society the Havik would then have been in an advantageous position through access to more advanced technology, including agriculture. Much of population growth in India over the millennia must have related to gradual agricultural colonization of new areas followed by intensification of agriculture through irrigation in areas already under cultivation (Jha 1974). Groups assigned low status, such as the Mukri, may have subsisted as hunter-gatherers prior to formation of the caste society.

In the hierarchical caste society the Brahmin (including the Havik) were granted land by rulers in order to establish irrigated tracts and temples. With control over irrigation, spice gardening, and priestly functions, they have been capable of generating a surplus for at least the past 2,000 years. During that period they were continually spreading to new territories and presumably growing in numbers. The lowest-ranking castes (including the Mukri) were landless and less likely to generate a surplus through such occupations characteristic of erstwhile untouchables as hunting, fishing, gathering, weaving of reed mats or bamboo baskets, tanning of leather, or farm labor. It is plausible that there was a gradient of population growth rates corresponding to social hierarchy: Brahmin groups continued to expand because of greater economic surplus, while scheduled castes maintained their population sizes or underwent multiple bottlenecks. Figure 8 indicates likely, simplified scenarios compatible with the pairwise-difference distributions described here.

Other cultural practices within the caste system also ensure that a Brahmin caste group grows at a higher rate than lower status caste groups. In the rare instances that marriage is outside the endogamous group, offspring belong to the father's group. Such marriages traditionally involved a man from a higher status group marrying a lower status woman (Powell 1896). Children would therefore be absorbed into the higher status group. This ensures that the higher status social groups grow in size at a somewhat faster rate, all other factors being equal.

The tree inferred from sequences for all the Indian samples reveals no clear separation among the subpopulations. A similar pattern, with no identifiable caste-specific clades, was recently reported by Bamshad et al. (1994) for a study of 40 individuals from four caste groups in Andrah Pradesh, India. Although the Havik and Mukri examined here are among the highest and lowest status castes, respectively, sequences from both samples are found scattered throughout the tree. Such a pattern might be evidence for some level of gene flow between these two subpopulations. A recent study indicates that prior to the 1980s ~3% of marriages were exogamous (Bhattacharyya 1993). This small amount of gene flow, if maintained over a number of generations, would certainly have an impact. In this region, however, no marriages between Mukri and Havik take place (Bhattacharyya 1993). In the hierarchical caste system the Havik and Mukri are separated by several other caste levels; marriage outside of a caste group only takes place between members of groups of roughly the same social status. The rate of gene flow between Havik and Mukri must therefore be orders of magnitude lower than 3%. While Havik and Mukri sequences are found together in each part of the tree shown in figure 3, Havik and Mukri sequences are generally separated by branches supported by bootstraps. This pattern indicates that very recent gene flow is unlikely. The lack of clustering according to caste affiliation is more likely an indication that the separation between mitochondrial DNA lineages predates the separation of populations (Avise et al. 1984).

Indian Sequences in a Broader Context

Previous genetic studies have demonstrated that Indian populations cluster most closely with west Asian populations, including Turks, Iranians, Caucasians, and Lebanese (Cavalli-Sforza et al. 1994). The Indians and west Asians form a Caucasoid cluster when compared with Asians from further north and east.

As do sequences of other populations, the Indian sequences, both Mukri and Havik, tend to scatter throughout the starlike M_2 cluster that includes Europeans, Asians, New Guineans, and some Africans (fig. 5). This pattern is somewhat surprising, considering that these individuals were sampled from a very small geographic area, and indicates that the common ancestor of these Indian lineages predates the divergence among Eurasian populations. That is, diverse mitochondrial lineages appear to have existed prior to the separation of these Indian populations and even prior to the separation of non-Africans into various ethnic groups. Furthermore,

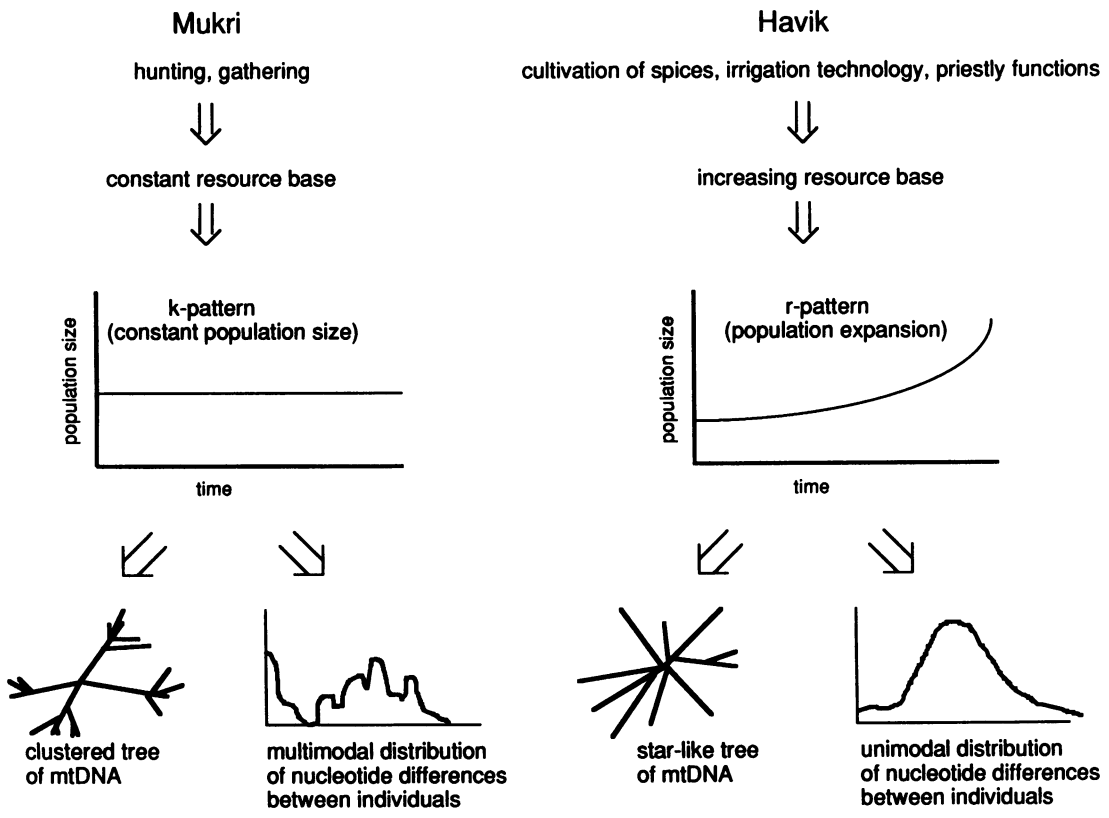

Figure 8 Illustration of possible scenarios leading to the difference between the trees and pairwise-difference distributions of the Mukri and Havik samples.

most clusters of Indian sequences join sequences of other populations near the center of this cluster, suggesting that little mixture between these Indian groups and the other populations sampled has taken place. The Indian populations considered here may have arisen as part of a broader Eurasian expansion. Since then the Havik appear to have been expanding, while the Mukri have either maintained a roughly constant population size or undergone multiple bottlenecks.

While the !Kung were also drawn from a very restricted geographic region, the two samples (!Kung and Indian) form very different patterns in the tree shown in figure 5. The !Kung form a single cluster that includes no individuals from other populations. This observation has implications for sampling of human populations; while sampling criteria may be consistently geographic or linguistic, widely different genetic patterns may emerge.

Major Human Expansions

Archaeological evidence suggests that the first migration of anatomically modern humans from Africa to other world regions took place roughly 100,000 years ago (Brauer 1989; Clark 1989). If population growth accompanied such migration, it did not leave extensive archaeological traces. Nonetheless, migration of modern or near-modern peoples from northern Africa into the Middle East appears to have taken place during the Last Interglaciation (Klein 1989, p. 341), and may have been accompanied by population growth. Harpending et al. (1993) suggest that many independent population expansions occurred following a period of isolation, after modern human groups initially migrated from Africa. We propose an alternate hypothesis: that a single, major expansion occurred after a period of isolation between African and non-African populations, i.e., no extended period of isolation among non-African groups took place. Assuming an expansion in Africa beginning 100,000 years ago, we estimated that such a non-African expansion may have taken place ~65,000 years ago. This date is roughly consistent with archaeological data indicating that modern humans had reached Australia by 50,000 or more years ago (Roberts et al. 1990).

If, as Klein (1994) has suggested, population growth in Africa did not occur until ≤60,000 years ago, the non-African expansion is estimated to have taken place closer to 40,000 years ago. In this case the arrival of the first modern humans in Australia would have preceded the proposed non-African expansion. The proposed expansion would roughly coincide with entry of modern humans into Europe (Brauer 1989).

Figure 6*c* shows the distribution of pairwise differ-

ences between African and non-African sequences. The mean of 13.04 differences, while not much greater than the mean for the within-African distribution (12.80 differences), is substantially greater than the mean for the non-African distribution (8.96 differences). Such a pattern suggests that Africans and non-Africans separated at about the time of an expansion. Furthermore, the starlike M_2 cluster in figure 5 includes not only non-Africans but also a number of African individuals, suggesting that the non-African population arose from a subpopulation in Africa.

The value of θ_1 for the non-African distribution, corresponding to the population size after the expansion, is essentially twice that of the African distribution (table 3). The θ_0 values, however, are roughly the same. The non-African distribution has a lower variance than the African distribution, reflecting a more starlike tree for these sequences. Such a pattern suggests that the non-African populations underwent a larger expansion than did the African populations. This is consistent with archaeological data showing little evidence of a major expansion within Africa (Klein 1989).

The world tree inferred here (fig. 5) is consistent with a major expansion of modern humans taking place outside of Africa some time after an initial expansion within Africa. This conclusion is comparable to dates estimated by other groups studying mtDNA data. Ballinger et al. (1992) suggest that an origin of Mongoloid people took place ~59,000–118,000 (average 88,950) years ago, while Lum et al. (1994) estimated a date of 85,000 years before present for the sequence divergence of Polynesians. One question that arises is where such an expansion might have originated, if it was centered in a single area. While it may have begun in or near Africa, it may also have begun further east. Ballinger et al (1992) suggest an origin in southern China. India, geographically central to Eurasia and Oceania, may have been relatively easy to reach from Africa along the coast and is another candidate for a center of origin. We hope to acquire further insights into the histories of these and other Indian groups, in the context of world populations, through analyses of additional mtDNA sequences and of nuclear DNA loci.

Acknowledgments

We would like to thank M. Stoneking for supplying computer files of mtDNA sequences; M. Slatkin, H. Harpending, and A. Rogers for helpful comments on an earlier version of the manuscript; A. Rogers for providing the "mom" computer program; F. Calafell for discussions and computations; and R. Scott for graphics. We also thank Anindya Sinha, J. Pillai, and Subash Chandran of the Centre for Ecological Sciences, Indian Institute of Science for their assistance with collection of samples. This research was supported by NIH grant GM 28428.

References

Aquadro CF, Greenberg BD (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics 103:287–312

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon ICE, et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

Avise JC, Neigel JE, Arnold J (1984) Demographic influences on mitochondrial DNA lineage survivorship in animal populations. J Mol Evol 20:99–105

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen K-H, et al (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. Genetics 130:139–152

Bamshad M, Jorde LB, Crawford MH, Cann RL (1994) Human mitochondrial DNA evolution in east Indian caste populations of Andrah Pradesh. Am J Phys Anthropol Suppl 18:53

Bhattacharyya S (1993) Ecological organization of Indian rural populations. PhD dissertation, Indian Institute of Science, Bangalore

Brauer G (1989) The evolution of modern humans: a comparison of the African and non-African evidence. In: Mellars P, Stringer C (eds), The human revolution: behavioural and biological perspectives on the origins of modern humans. Edinburgh University Press, Edinburgh, pp 123–154

Brown WM, George M Jr, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci USA 76:1967–1971

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31–35

Cavalli-Sforza LL, Piazza A, Menozzi P (1994) History and geography of human genes. Princeton University Press, Princeton

Clark JD (1989) The origin and spread of modern humans: a broad perspective on the African evidence. In: Mellars P, Stringer C (eds) The human revolution: behavioural and biological perspectives on the origins of modern humans. Edinburgh University Press, Edinburgh, pp 565–588

D'Aquila RT, Bechtel LJ, Videler JA, Eron JJ, Gorczyca P, Kaplan JC (1991) Maximising sensitivity and specificity of PCR by pre-amplification heating. Nucleic Acids Res 19:3749

Efron B (1982) The jackknife, bootstrap, and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

——— (1989) PHYLIP-Phylogeny inference package (version 3.2). Cladistics 5:164–166

Gadgil M, Guha R (1993) This fissured land: an ecological history of India. University of California Press, Berkeley

Gadgil M, Iyer P (1989) On the diversification of common property resource use by Indian society. In: Berkes F (ed) Common property resources: ecology and community-based sustainable development. Belhaven Press, London, pp 240–255

Giles RE, Blanc H, Cann HM, Wallace DC (1980) Maternal inheritance of mitochondrial DNA. Proc Natl Acad Sci USA 77:6715–6719

Gyllensten U, Wharton D, Josefsson A, Wilson AC (1991) Paternal inheritance of mitochondrial DNA in mice. Nature 352:255–257

Harpending HC, Sherry ST, Rogers AR, Stoneking M (1993) The genetic structure of ancient human populations. Curr Anthropol 34:483–496

Hartl DL, Clark AG (1989) Principles of population genetics. Sinauer, Sunderland, MA

Herre W, Rohrs M (1977) Zoological considerations of farming and domestication. In: Reed C (ed) Origins of agriculture. Manton, The Hague, pp 245–280

Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73:237–244

Higuchi R (1989) Simple and rapid preparation of samples for PCR. In: Erlich HA (ed) PCR technology: principles and applications for DNA amplification. Stockton, New York, pp 31–38

Hoeh WR, Blakley KH, Brown WM (1991) Heteroplasmy suggests limited biparental inheritance of Mytilus mitochondrial DNA. Science 251:1488–1451

Jha DN (1974) Temples as landed magnates in early medieval south India. In: Sharma RS, Jha V (eds) Indian society: historical probings. Manohar, New Delhi

Joshi NV, Gadghil M, Patil S (1993) Exploring cultural diversity of the people of India. Curr Sci 64:10–17

Karve I (1961) Hindu society: an interpretation. Sangam, Pune, India

Karve I, Malhotra KC (1968) A biological comparison of eight endogamous groups of the same rank. Curr Anthropol 9:109–124

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies nucleotide sequences. J Mol Evol 16:111–120

Klein RG (1989) The human career: human biological and cultural origins. University of Chicago Press, Chicago

——— (1994) An introduction to the problem of modern human origins. In: Nitecki MH, Nitecki V (eds) Origins of anatomically modern humans, Plenum, New York, pp 3–17

Lum JK, Rickards O, Ching C, Cann RL (1994) Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. Hum Biol 66:567–590

Malhotra KC (1984) Population structure among the Dhangar caste-cluster of Maharashtra, India. In: Lukacs JR (ed) The people of south Asia: the biological anthropology of India, Pakistan, and Nepal. Plenum, New York, pp 295–324

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Piazza A, Menozzi P, Cavalli-Sforza L (1981) The making and testing of geographic gene-frequency maps. Biometrics 37:635–659

Powell B (1896; reprint 1981) Indian village community. Modern Asian Publishers, Delhi

Roberts RG, Jones R, Smith MA (1990) Thermoluminescence dating of a 50,000-year-old human occupation site in northern Australia. Nature 345:153–156

Robin ED, Wong R (1988) Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J Cell Physiol 136:507–513

Rogers AR. Genetic evidence for a Pleistocene population explosion. Evolution (in press)

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

Roychoudhury AK (1984) Genetic relationships between Indian populations and their neighbors. In: Lukacs JR (ed) The people of south Asia: the biological anthropology of India, Pakistan, and Nepal. Plenum, New York, pp 283–293

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129:555–562

Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. Univ Kansas Sci Bull 38:1409–1437

Vidyarthi LP (1983) Tribes of India. In: Satyavati GV (ed) Peoples of India: some genetical aspects. Indian Council of Medical Research, New Delhi, pp 85–103

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507

Wakeley J (1993) Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. J Mol Evol 37:613–623

Wilson AC, Cann RL, Carr SM, George M, Gyllensten UB, Helm-Bychowski KM, Higuchi RG, et al (1985) Mitochondrial DNA and two perspectives in evolutionary genetics. Biol J Linnean Soc 26:375–400

Wolfram S (1988) Mathematica: a system for doing mathematics by computer. Addison-Wesley, Menlo Park