Project MUSE®

Genetic Diversity and Relationships Among the Tribes of Meghalaya Compared to Other Indian and Continental Populations

B.T. LANGSTIEH,[1,2] B. MOHAN REDDY,[2] K. THANGARAJ,[3] V. KUMAR,[2] AND LALJI SINGH[3]

Abstract The autosomal AmpFLSTR markers validated and widely used for forensic applications are used in this study to examine the extent of diversity and genetic relationships among nine Meghalaya populations. Altogether, 932 chromosomes from 9 populations were analyzed using 9 tetrameric AmpFLSTR loci. The included populations were all seven subtribes of the Austro-Asiatic Mon-Khmer-speaking Khasi and the neighboring Tibeto-Burman Garo. The Lyngngam, which are linguistically closer to the Khasi but are culturally intermediate between the Khasi and the Garo, are also included in the study. Although most of the microsatellite loci are highly polymorphic in each of these populations, the allele distributions are fairly uniform across the Meghalaya populations, suggesting relative homogeneity among them. Concurrent with this, the coefficient of gene differentiation (G_{ST}) is observed to be low (0.026 ± 0.002). This is naturally reflected in the lack of clear differentiation and clustering pattern of the Meghalaya tribes based on either geographic proximity or the historical or current affiliations of these tribes. Analysis of molecular variance (AMOVA) suggests no significant population structure. The structure analysis further suggests that, barring War-Khasi and Pnar, no other population shows any semblance of genetic identity. Even the position of the linguistically distinct Garo is not portrayed as separate from the Khasi. However, when comparable data from other Indian, Southeast Asian, and other continental populations were analyzed, the Meghalaya populations formed a compact cluster clearly separated from other populations, suggesting genetic identity of the Meghalaya populations as a whole. These results are concurrent with the hypothesis of a common and recent origin of these Meghalaya populations, whose genetic differentiation is overwhelmed by the homogenizing effect of continuous gene flow.

[1]Department of Anthropology, North Eastern Hill University, Shillong, Meghalaya, India.
[2]Anthropology and Human Genetics Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Kolkata, India.
[3]Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad, India.

Human Biology, August 2004, v. 76, no. 4, pp. 569–590.
Copyright © 2004 Wayne State University Press, Detroit, Michigan 48201-1309

KEY WORDS: KHASI SUBTRIBES, GARO POPULATION, LYNGNGAM POPULATION, MEGHALAYA POPULATIONS, INDIA, AMPFLSTR LOCI, D3S1358, D8S1179, D5S181, VWA, D21S11, D13S317, FGA, D7S820, D18S5, GENE FLOW, GENETIC HOMOGENIZATION.

The northeastern part of India is referred to as a melting pot of Mongoloid, Australoid, and Caucasoid populations, which is exhibited in the unique sociocultural diversity of the region. It has been described as the corridor for the influx of migratory populations from Southeast Asia and neighboring Tibet, Myanmar (Burma), Thailand, etc. These populations might have settled in this region at different times and probably arrived in different batches as hordes of food gatherers, hunters, and warriors, according to prehistoric archeological evidence from the Garo hills, Meghalaya, that suggests that this region might have been inhabited as early as the Paleolithic period (Sharma 1966, 1980; Hussain 1991).

There are two indigenous and predominant tribal clusters, namely, the Khasi and the Garo, perhaps one of the few populations in India and the world that follow the system of matrilineal descent and matrilocal residence. Although the Khasi, who occupy the central and eastern regions of Meghalaya, are the only Mon-Khmer Austro-Asiatic speakers located as a pocket amid the ethnic majority of Asian populations of Sino-Tibeto-Burman origin in the entire northeastern region, the Garo, who inhabit the western parts of Meghalaya, are Tibeto-Burman speakers. Given the hypothesis that northeast India served as a major corridor for early human migrations into and out of the Indian subcontinent and particularly because of the presence of the Austro-Asiatic Khasi, the study of populations from this region assumes special significance in answering questions about the peopling of India and routes of migration. Some recent studies involving molecular genetic markers (Basu et al. 2003) speculate that Austro-Asiatic tribes were the earliest settlers in India.

Comprehensive genome diversity studies of different ethnic, regional, and linguistic groups of India are needed to find unequivocal answers to some of the issues concerning the history and peopling of this region and also to test some of the current anthropological hypotheses. This need prompted us to initiate studies among different regional and linguistic populations of India. These samples are being analyzed for different sets of DNA markers: mitochondrial, Y chromosome based, and autosomal. Here, we report findings based on the analyses of nine amplified-fragment-length short-tandem-repeat (AmpFLSTR) loci among the tribal populations of Meghalaya in the northeastern part of India. We examine the nature and extent of genetic diversity and relationships among the nine tribal populations of Meghalaya, which represent both the linguistic and the geographic heterogeneity of the state in relation to other Indian and continental populations. Further, we probe whether the matrilineal system prevalent among these populations has a role in the observed pattern of diversity and relationships among the Meghalaya populations, thus exploring the influence of social system on the genetic structure of tribal populations of India.

Microsatellite loci have been widely used to study genetic relationships among human populations on the continental (Bowcock et al. 1994; Cavalli-Sforza et al. 1994; Deka et al. 1995a, 1995b; Nei and Takezaki 1996; Perez-Lezaun et al. 1997; Jorde et al. 1997; Eller 1999), regional (Parra et al. 1999;

Table 1. Sample Sizes, Location of Study, and Linguistic Background of the Nine Tribal Populations of Meghalaya

Population	*Approximate Size*	*Sample Size*	*Traditional Occupation*	*Distribution in Meghalaya*	*Linguistic Family*
Lyngngam	6,000	156	Shifting cultivators	West Khasi Hills District	Austro-Asiatic (Mon-Khmer)
Nongtrai	6,000	90	Shifting cultivators	West Khasi Hills District	Austro-Asiatic (Mon-Khmer)
Maram	200,000	96	Settled agriculturists	West Khasi Hills District	Austro-Asiatic (Mon-Khmer)
Khynriam	550,548	146	Settled agriculturists	East Khasi Hills District	Austro-Asiatic (Mon-Khmer)
Pnar	259,667	100	Settled agriculturists	Jaintia Hills District	Austro-Asiatic (Mon-Khmer)
War Khasi	33,000	80	Horticulturists	East Khasi Hills District	Austro-Asiatic (Mon-Khmer)
War Jaintia	36,025	46	Horticulturists	Jaintia Hills District	Austro-Asiatic (Mon-Khmer)
Bhoi	179,630	90	Shifting cultivators	Ri-Bhoi District	Austro-Asiatic (Mon-Khmer)
Garo	710,757	128	Shifting cultivators and settled agriculturists	South Garo Hill District and others	Tibeto-Burman

Reddy et al. 2001a; Dutta et al. 2002), and local levels (Reddy et al. 2001c). These studies yielded results that were usually consistent with the expected patterns on the basis of ethnohistoric, linguistic, and/or geographic backgrounds. A general concordance with the trees inferred from other types of nuclear markers, including classical genetic markers, RFLPs and *Alu* insertion polymorphisms, has also been observed (Nei and Takezaki 1996; Stoneking et al. 1997). However, simulation results indicate that the microsatellite loci generally provide a more reliable picture of the phylogenetic relationship of closely related populations than of distantly related ones (Nei and Takezaki 1996). On the basis of 13 STR loci, we have recently observed that these markers help to reconstruct the short evolutionary history at the level of subcastes of an Indian caste, because these loci seem to have left signatures of subcaste endogamy (Reddy et al. 2001c). It remains to be seen, however, whether similar patterns can be found in substructured tribal populations of India.

Materials and Methods

Background of the Studied Populations. For the present study we have considered seven Khasi subtribes in addition to the Lyngngam and the Garo. The names of the different tribes, their geographic distribution, population size, etc. are furnished in Table 1 and Figure 1. The Khasi and the Garo are two independent tribes with distinct origins and languages, although they are similar in the

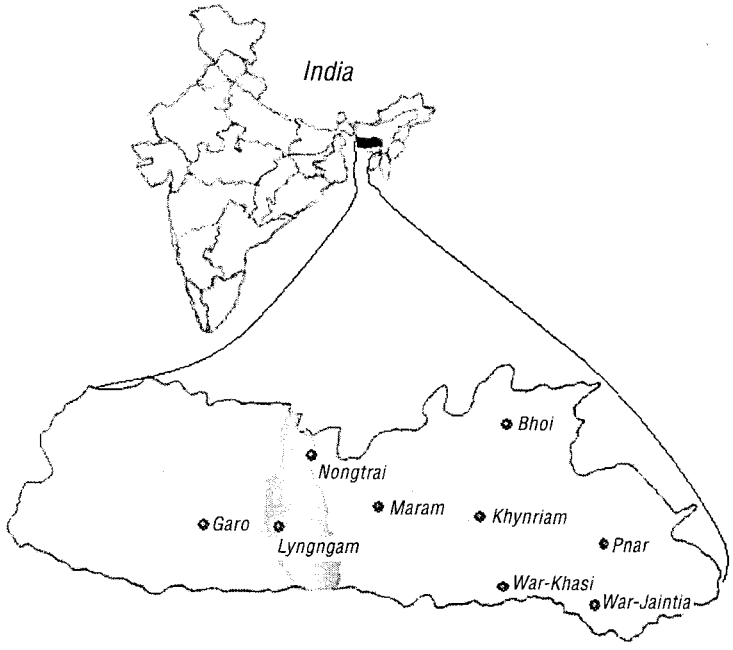

Figure 1. Map of Meghalaya showing the core areas of distribution of different subtribes of the Khasi and the Garo.

practice of the matrilineal rule of descent and matrilocal residence. Interestingly, the Garo of Meghalaya are apparently the only group of Tibeto-Burman speakers who practice matriarchy. Each of these tribes has a number of subdivisions, structured on the geographic and dialectal identity. The seven Khasi tribes live in the Khasi and Jaintia hills, and the Garo live in the Garo hills of Meghalaya; the two groups' geographic distributions do not overlap. Distribution of particularly the Lyngngam in the West Khasi hills is in close proximity and contiguous to the Garo hills inhabited by the Garo. As a result, although the Khasi refer to the Lyngngam as Lyngngam or Langám and consider them as belonging to one of their own groups, the Garo call them Megam and treat them as part of one of the 12 subtribes of the Garo. Their ethnic identity therefore is in dispute, although their language is more akin to the Khasi's.

The history of migration of the Garo is relatively recent and originated from Torua, an area in Tibet that borders China. The Austro-Asiatic linguistic family is a large group of Austric speakers most commonly found in Southeast Asia (Cambodia, Vietnam, Laos, Burma, etc.). They are represented in India by the Mundari and Mon-Khmer linguistic subfamilies. Although the Mundari form of Austro-Asiatic languages is widely spoken by certain indigenous tribal populations of central and eastern India, besides the Khasi only the Shompen and Nicobarese of the Andaman and Nicobar Islands speak Mon-Khmer languages.

Sampling and DNA Isolation. Intravenous blood samples (3–5 ml) were collected from 466 individuals belonging to 9 subgroups of the Meghalaya tribes. Participants gave their informed written consent. The tribewise sample distribution is given in Table 1. DNA was isolated from the samples following standard protocols. Erythrocytes were lysed with 15.0 ml of EL buffer (10 m*M* Tris, pH 8.0; 320 m*M* sucrose; 5 m*M* $MgCl_2$; and 1% Triton X-100) for 5 min. After complete lysis of erythrocytes, leukocytes were pelleted by centrifugation at 1,500 rpm for 5 min. The leukocyte pellet was dissolved in 8.0 ml of LL buffer (400 m*M* Tris, 60 m*M* EDTA, 150 m*M* NaCl, and 1% SDS) and mixed thoroughly. To this, 2.0 ml of 5 *M* sodium perchlorate was added and mixed thoroughly for 2–3 min. DNA was precipitated after extracting once with phenol:chloroform and once with chloroform; DNA was washed with 70% ethanol and dissolved in TE buffer (pH 8.0). The extracted DNA was quantified using a spectrophotometer method followed by quantification in 0.8% agarose gel (Maniatis et al. 1989).

STR Profiling. Nine STR loci were amplified using the AmpFLSTR Profiler Plus kit (ABI, Applied Biosystems), according to the manufacturer's instructions. Amplified samples were analyzed in an ABI 377 automated DNA sequencer, and the GeneScan and Genotyper softwares (Perkin Elmer) were used to obtain the allele designations at the D3S1358, D8S1179, D5S181, VWA, D21S11, D13S317, FGA, D7S820, and D18S5 loci.

Statistical Methods. The allele frequencies were estimated using the direct gene-counting procedure. Average heterozygosity and the coefficient of gene differentiation (G_{ST}) along with their standard errors were obtained following the method of Nei (1987). The genetic distances were computed using the modified Cavalli-Sforza distance (D_A) of Nei et al. (1983). The distances were also obtained using the stepwise weighted genetic distance measure (D_{SW}) of Shriver et al. (1995) and Reynolds's ϕ_{ST} or F_{ST} (Reynolds et al. 1983). Although D_A is not linear with evolutionary time, it is efficient in obtaining the most correct phylogenetic relationships among closely related populations (Takezaki and Nei 1996). On the other hand, F_{ST} or ϕ_{ST} (Reynolds et al. 1983) is a modified form of Cavalli-Sforza's chord distance, with the assumption that there is no new mutation and that all gene frequency changes are due to genetic drift. Constant and equal population sizes were not assumed. Therefore F_{ST} or ϕ_{ST} may be the most appropriate measure of genetic distance for the populations of Meghalaya with short evolutionary history.

The neighbor-joining algorithm (Saitou and Nei 1987) was used to construct the phylogenetic trees. Computations were performed using the NJBAFD program (supplied by N. Takezaki, National Institute of Genetics, Mishima, Japan) and PHYLIP, version 3.573. The congruence between the geographic distance matrix and the genetic distance matrix D_A was tested using the Mantel software (Relethford 1993). Finally, multidimensional scaling (MDS) plots of

the populations on the basis of genetic distances were obtained with the help of SPSS, version 7. Hardy–Weinberg equilibrium (Guo and Thompson 1992) and population genetic structure as inferred by the analysis of molecular variance (AMOVA) (Excoffier et al. 1992) were performed using the Arlequin software, version 2.00 (Schneider et al. 1997; available at http://www.anthropologie.unige.ch/arlequin).

The regression model of Harpending and Ward (1982) was used to study the possible effects of genetic drift and admixture on the substructured Meghalaya populations. According to this model, the average heterozygosity of the *i*th population (H_i) should be equal to the overall mean heterozygosity of the entire population (in this case, Meghalaya populations), H_t, multiplied by $(1 - r_{ii})$, where r_{ii} is the genetic distance of a particular population from the gene frequency centroid. If gene flow from outside the region varies in amount from population to population, then this linear relationship no longer holds. Isolated groups will be less heterozygous than the linear prediction and hence will lie below the expected regression line, whereas populations receiving more gene flow from outside will be more heterozygous and therefore will lie above the regression line.

We also performed a Markov chain Monte Carlo analysis of population structure using the software Structure, version 2 (available at http://pritch.bsd.uchicago.edu), which implements a model-based clustering method for grouping individual populations and for identifying migrants and admixed individuals. This approach assumes a model with K populations, each of which is characterized by a set of allele frequencies at each locus; individuals are assigned to one population or jointly to two or more populations if their genotypes indicate that they are admixed. Individuals are grouped into populations in such a way as to achieve Hardy–Weinberg equilibrium and linkage equilibrium (Pritchard et al. 2000). Because the populations analyzed in this study were closely related in their linguistic affiliations (except the Garo) and geographic distribution, we used the improved model of allele frequency correlations as elucidated by Falush et al. (2003). We also used the model with a priori population information and without. Furthermore, we performed the analysis with different combinations of probability that an individual is an immigrant to a population in recent generations. For the analysis the burn-in lengths were chosen so that the summary statistics would converge, and the simulation run was quite long (8×10^6) to get accurate estimates of parameters.

Results

Allele Frequency Distributions. Allele frequency distributions at the nine STR loci are given in Appendix 1. All the loci were highly polymorphic among the Meghalaya populations, with the number of alleles ranging from 8 for D13S317 to as many as 33 for D21S11. However, when the population-specific

allele range was considered, a minimum of 5 alleles was found at D3S1358 and D5S818 in the War Jaintia population and a maximum of 23 alleles was observed at D21S11 in the Tibeto-Burman Garo. Overall, D21S11, FGA, and D18S51 showed the largest number of alleles, and D3S1358 and D5S818 showed the smallest number of alleles. The extent of heterogeneity observed both within and between the populations concurs with this trend in the number of alleles. However, the shape of the distribution of allele frequencies is fairly uniform across most populations for most of these loci, except again for the few most polymorphic loci. The predominant alleles at each locus remained within roughly the same range of repeats across different populations. The exact test (Guo and Thompson 1992) for Hardy–Weinberg equilibrium suggested significant departures for several locus–population combinations, and more often these departures were due to deficiency of observed heterozygosity, although in a few cases it was found to be due to greater observed heterozygosity.

Genetic Diversity Within and Between Meghalaya Populations. Locus-wise and average heterozygosity depicting within-population heterogeneity are presented in Table 2 along with the values of G_{ST} for each locus and for the average for the nine loci. Although D3S1358, D5S818, and VWA show relatively low heterozygosity within populations and a low coefficient of gene differentiation (G_{ST}) among populations, it is at the FGA, D21S11, and D7S820 loci, in that order, that one finds relatively greater heterozygosity and genetic differentiation; this concurs with the observed number of alleles. Except in the War Khasi ($H = 0.794$), the average heterozygosity is uniformly high among the Meghalaya populations, in a narrow range of 0.826–0.842. However, the average coefficient of gene differentiation is relatively low at 0.026 ± 0.002 and varies between 0.016 at D3S1358, D5S818, and VWA and 0.04 at FGA.

The AMOVA considering the nine Meghalaya tribes grouped into four groups based on geographic proximity and/or ethnohistoric association suggests little and insignificant variation (0.13%) among the groups (Table 3). Further, the variation among the populations within the groups is rather small (1.67%), although statistically significant. Thus most of the variation is found between individuals within the populations, reiterating the relative homogeneity of the Meghalaya populations as a whole.

Genetic Relationships Among Meghalaya Populations. Given the similarity in the trees based on D_A and F_{ST} and given the greater suitability of D_A for populations with relatively short evolutionary history, we present the neighbor-joining tree (Figure 2) based on only D_A, although we have constructed trees based on F_{ST}, D_{SW}, and D_{MYU} distances as well. The following salient features of the findings emerge: In the trees based on D_A and F_{ST} the War subgroups (War Jaintia and War Khasi), which are contiguously distributed in southern Meghalaya, form a distinct subcluster. The Khynriam, which is another subgroup of the Khasi that inhabits an area close to both War subgroups, separates from the main

Table 2. Locus, Average Heterozygosity and G_{ST} Values for the Nine AmpFLSTR Profiler Plus Loci for the Tribal Populations of Meghalaya

Locus	*Lyngngam*	*Garo*	*Nongtrai*	*Maram*	*Khynriam*	*Pnar*	*Bhoi*	*War Khasi*	*War Jaintia*	G_{ST}
D3S1358	0.700	0.706	0.747	0.743	0.752	0.773	0.747	0.750	0.738	0.016
D8S1179	0.863	0.867	0.867	0.874	0.854	0.867	0.826	0.847	0.855	0.023
D5S818	0.781	0.790	0.771	0.739	0.770	0.729	0.775	0.777	0.773	0.016
VWA	0.833	0.806	0.790	0.855	0.797	0.788	0.809	0.784	0.799	0.016
D21S11	0.928	0.915	0.898	0.896	0.879	0.902	0.924	0.857	0.899	0.035
D13S317	0.826	0.853	0.853	0.832	0.838	0.844	0.837	0.766	0.816	0.029
FGA	0.914	0.933	0.920	0.888	0.932	0.914	0.927	0.843	0.925	0.040
D7S820	0.806	0.811	0.799	0.802	0.815	0.803	0.838	0.805	0.839	0.030
D18S51	0.871	0.892	0.789	0.898	0.889	0.853	0.873	0.721	0.849	0.026
Average	0.836	0.842	0.826	0.836	0.836	0.830	0.840	0.794	0.833	0.026 ± 0.002

Table 3. AMOVA Results Based on the Population Structure of the Meghalaya Populations

Source of Variation	*df*	*Sum of Squares*	*Variance Components*	*Percentage of Variation*
Among groups	3	35.48	0.00506	0.13
Among populations within groups	5	50.23	0.06380	1.67
Within populations	925	3462.94	3.74372	98.19
Total	933	3548.65	3.81258	

Group 1 = Bhoi and Garo; Group 2 = Lyngngam, Nongtrai, and Maram; Group 3 = Khynriam and War Khasi; Group 4 = Pnar and War Jaintia.

cluster constituting the remaining six Meghalaya populations and joins the War cluster as an independent element. This cluster has the highest proportion of bootstraps (65%) and branches out first from the rest, suggesting greater antiquity. We note that the Amwi dialect spoken by these people is believed to be the closest to the Mon-Khmer groups, although no evidence is available at present to substantiate this belief. In the second cluster the internal branches generally show low bootstrap values (34–57%), and the population subclusters do not follow any discernible pattern. Overall, neither the neighbor-joining tree based on D_A nor the tree based on F_{ST} could separate the linguistically distinct Garo from the Khasi populations. Nor was there any clear-cut pattern of clustering, based on any rational criteria, apparent in the Khasi subpopulations.

The limitations inherent in the trees resulting from the imposition of bifurcation topology can be partially surmounted by using other graphical methods, such as principal components analyses or multidimensional scaling, to represent genetic relationships of the populations. Therefore we present a two-dimensional plot of the Meghalaya populations based on multidimensional scaling of the D_A and F_{ST} distance matrices (Figure 3). The two plots appear to be mirror images

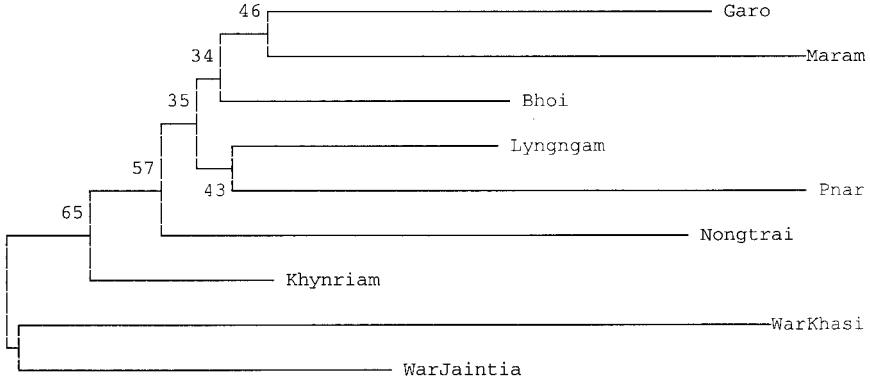

Figure 2. Neighbor-joining tree depicting the genetic relationships among the Meghalaya tribes (based on Nei's D_A computed for the nine STR loci).

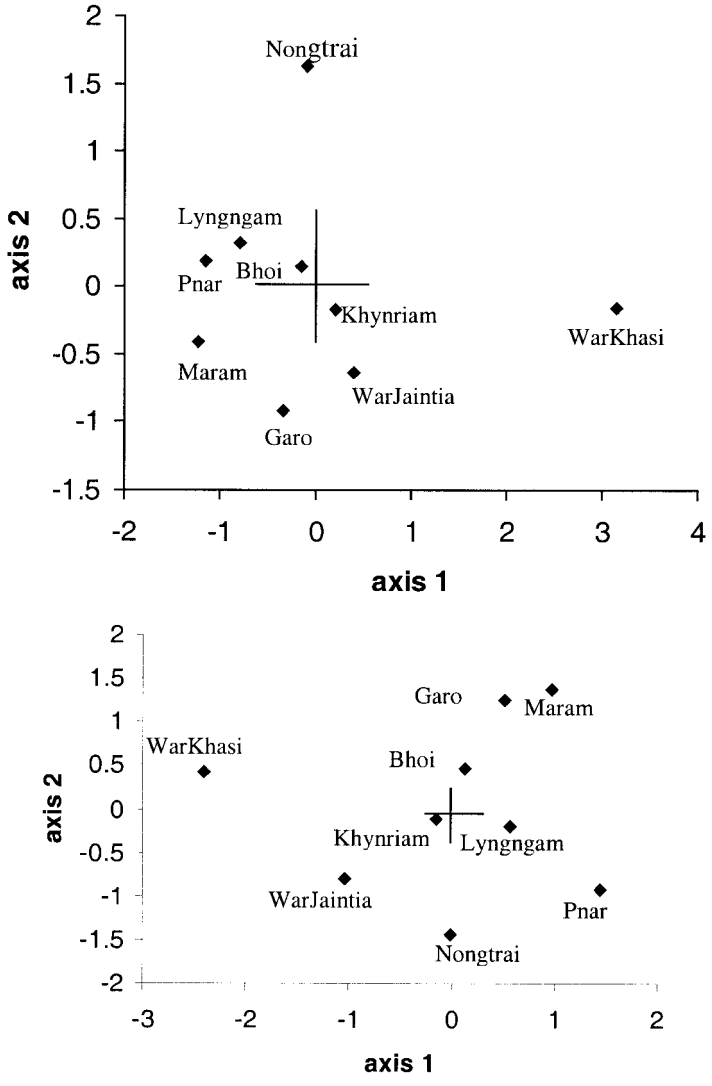

Figure 3. Projection of Meghalaya tribal populations on the two-dimensional space based on the multidimensional scaling of (top) the F_{ST} distances and (bottom) D_A distances derived from the profiler loci.

of each other somewhat. In the plot based on F_{ST} the differentiation appears to be more on the first axis, separating the War Khasi from the rest. On the second axis the Nongtrai and the Garo appear to be placed at the two extremes, although the Nongtrai appear to be relatively more divergent. The mirror image of this population configuration is largely true for the plot based on D_A. Overall, except for the War Khasi, the Meghalaya populations seem to show little divergence

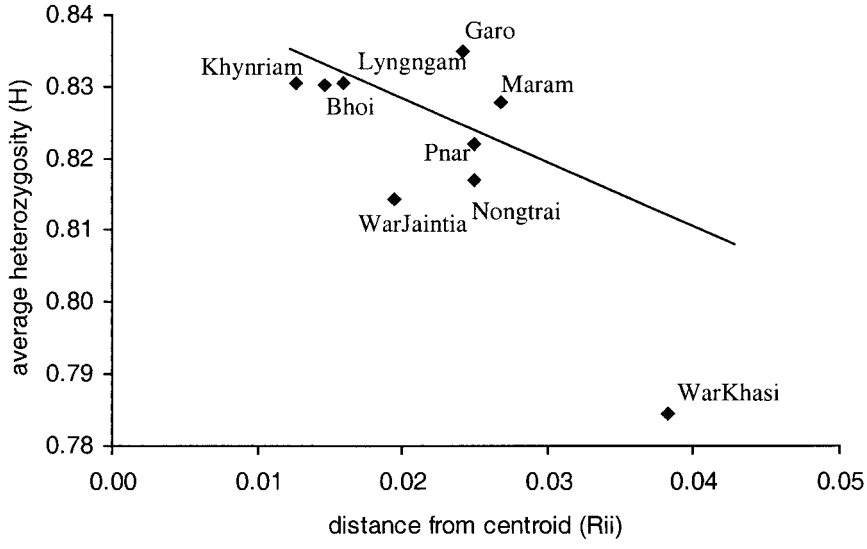

Figure 4. Regression plot of average heterozygosity (H) versus distance from centroid (r_{ii}) values of the Lyngngam and other Meghalaya populations, with the theoretical regression line based on nine loci autosomal STRs.

among them and hence no rational formation of population constellations in the multivariate space results. Although certain geographically contiguous groups tend to be placed together in the neighbor-joining plot and/or in the MDS plot, this is not consistent for other groups, resulting in low Mantel correlation between the geographic distance matrix and the genetic distance matrix ($r = 0.18$).

Regression of Average Heterozygosity on r_{ii}. The War Khasi, who have the largest r_{ii} and the lowest heterozygosity, are placed as an extreme outlier in the regression plot (Figure 4), probably reflecting the effect of long reproductive isolation and genetic drift. Given the geographic placement of the War Khasi in the more inaccessible and isolated locations, their position as an outlier below the regression line is expected, in contrast to the War Jaintia, who inhabit areas closer to the Pnar and to neighboring Bangladesh. The Lyngngam, like all the other Khasi populations, are placed near the regression line, suggesting no major effect of gene flow from outside these populations. Only the Maram and the Garo show some semblance of the effect of external gene flow, because they are placed above the theoretical line, albeit not as glaring outliers. These two populations had been exposed to intense outside contact as a result of the timber trade and coal mining for the last two decades, and this may be reflected by their position in the plot.

Structure Analysis. From the analysis of the populations without a priori information using an admixture model and repeating the run for different values

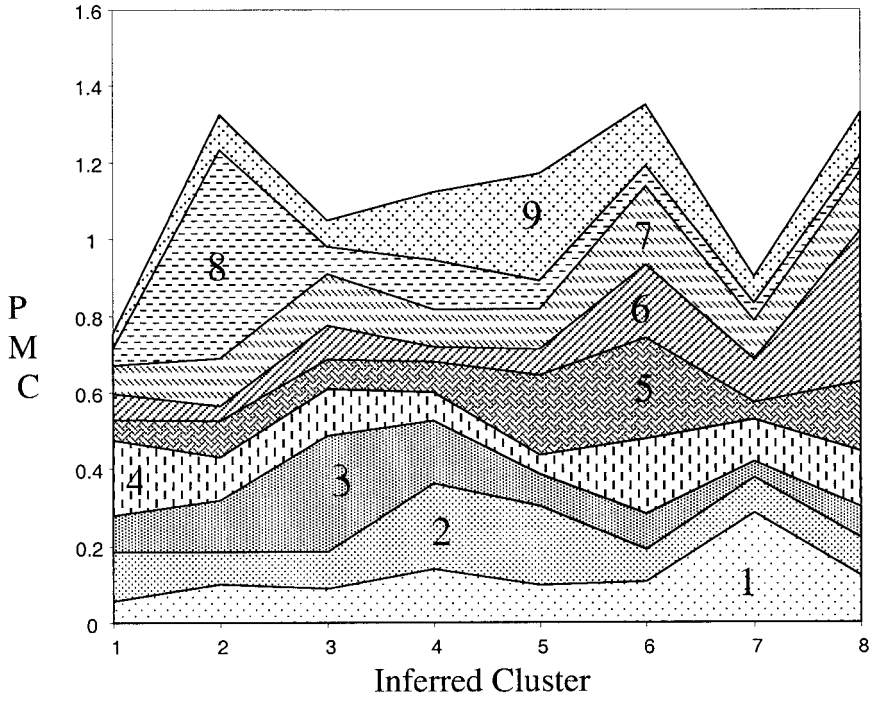

Figure 5. Proportion of membership coefficients (PMC) of the nine Meghalaya populations into the eight inferred clusters. The population(s) with the maximum membership coefficient in each cluster is indicated by its number. Populations (numbers inside figure): (1) Lyngngam, (2) Garo, (3) Nongtrai, (4) Maram, (5) Khynriam, (6) Pnar, (7) Bhoi, (8) War Khasi, (9) War Jaintia.

of K ($K = 3, \ldots, 10$), we found that, although the estimates of P_r (genotypic data $|$ K) increase with increasing value of K, the difference was not much for $K = 8$ and $K = 9$. Therefore we have chosen $K = 8$ (i.e., the number of putative populations to which individuals can be assigned).

Figure 5 is based on the proportion of membership of each predefined population in each of the clusters; it suggests that the maximum number of individuals of the Lyngngam, Garo, Maram, War Khasi, Pnar, War Jaintia, Khynriam, and Nongtrai is found in clusters 7, 4, 1, 2, 8, 5, 6, and 3, respectively. However, except for the War Khasi (54.1%) and the Pnar (39.5%), most of the others populations show relatively low values of the membership coefficient (19.4–30.1%). The individuals of Maram show uniform presence in most of the clusters; this is consistent with their central geographic position. A similar pattern is also depicted by the Bhoi, and these two populations therefore show by far the lowest values of the membership coefficient in any particular cluster. The triangle plot (figure not shown) suggests that there is indeed a great amount of overlap of individual members of different populations, reflecting probably continuous gene flow. Only the War Khasi and the Pnar show a semblance of population identities.

Table 4. Average G_{ST} Values for the Meghalaya Populations Compared to Other Populations in an Increasing Hierarchy of Populations

	Meghalaya	*Northeast India*	*All India*	*Southeast Asia*	*World*
G_{ST} ± S.E.	0.026 ± 0.002	0.046 ± 0.004	0.050 ± 0.004	0.046 ± 0.003	0.0 50 ± 0.004

A critical appraisal of the results of the membership coefficients suggests a structure with the Khynriam, Bhoi, Maram, and Pnar on the one hand and the War Jaintia, Garo, Lyngngam, and Maram on the other hand forming loose genetic conglomerations. However, when we use $K = 4$, a relatively better structure with three distinct clusters emerges: the War Khasi and War Jaintia in cluster 1, the Garo and the Nongtrai in cluster 2, and the Khynriam, Bhoi, Maram, and Pnar in cluster 3. All these populations have the largest membership coefficients in their respective clusters. Interestingly, the Lyngngam are the only population with the highest membership coefficient unique to cluster 4 and with relatively uniform membership in each of the four clusters.

Differentiation of Meghalaya Populations Compared to Other Indian and Continental Populations. To evaluate the extent of genetic differentiation in the populations at different levels of the hierarchy, we gathered comparable allele frequency data on 38 other populations. Nine of these populations are Indian, and the rest are from different parts of the world (see Figure 6 for population details). The G_{ST} values and their standard errors are given in Table 4, including populations from different levels in the hierarchy Meghalaya → northeast India- → all India → East and Southeast Asia → world. Consistent with increasing level in the hierarchy or complexity of the populations, the G_{ST} values systematically increase (0.026 in Meghalaya populations to 0.05 among the Indian populations). Beyond this the G_{ST} values fail to show any increase either when Asian or Southeast Asian populations ($G_{ST} = 0.046$) are included or when all the worldwide populations are included (0.05).

The neighbor-joining tree drawn on the basis of D_A distances (Figure 6) brings out six major and distinct clusters of populations: (1) all the tribal populations of Meghalaya; (2) all the Asian groups from East and Southeast Asia; (3) all native Americans; (4) certain populations of eastern and northeastern India consisting of two subclades, one formed by two upper castes from Bengal and the Meitei and Muslim groups from Manipur and the other formed by four Asian tribal populations from Manipur and north Bengal; (5) all the populations with African origins; and (6) all the European-derived populations from Europe and America. The most interesting observation is that all the populations of the present study form a single and compact cluster, which is a subclade of the major cluster of populations of East Asia and Southeast Asia. Overall, the neighbor-joining tree suggests emphatically that the nine tetranucleotide STRs used in this study amply demonstrate their utility in differentiating populations at a relatively

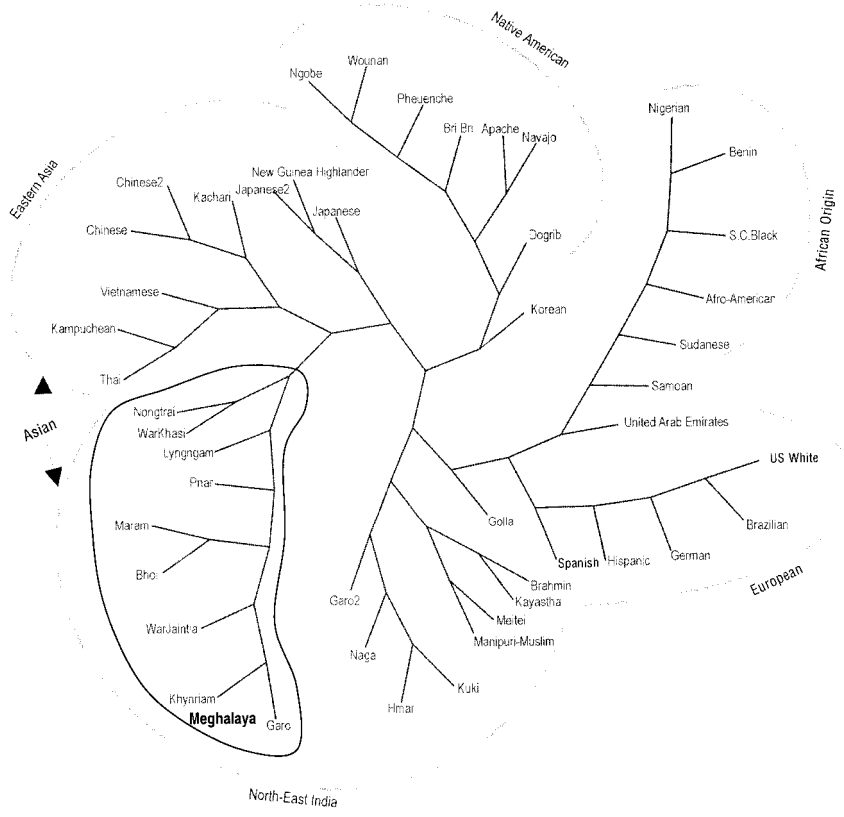

Figure 6. Neighbor-joining tree depicting the relationship of the Meghalaya tribes to other Indian and continental populations. Sources of comparative data: African, white, and Hispanic Americans (*AmpFLSTR Profiler Plus* 1998); Vietnamese (Borys et al. 1999a); Japanese (Borys et al. 1999b); Chinese (Fung et al. 2001); Golla (Reddy et al. 2001b); Garo2, Naga, Kuki, and Hmar (Chattopadhyay et al. 2001); Brahmin, Kayastha, Muslim (Manipur), and Meitei (Dutta et al. 2001); Sudanese, Nigerian, Benin, South Carolina Blacks, German, Spanish, United Arab Emirates, Brazilian, Chinese2, Japanese2, Kachari, Thai, Kampuchean, Dogrib, Ngöbe, Wounan, Bri Bri, Pehuenche, Samoan, and Papua New Guinea Highlanders (Sun et al. 2003).

higher level in the hierarchy, probably not at the lower levels, as, for example, in the case of the Meghalaya tribes.

Discussion

The use of AmpFLSTR profiler loci has been mostly limited to application in forensic investigations or human identification. To the best of our knowledge, only one other attempt has been made to study the patterns of variation using profiler loci among Indian populations (Dutta et al. 2002). In this study Dutta et

al. (2002) nevertheless succeeded in clearly portraying the genetic relationships expected on the basis of ethnohistoric, linguistic, and geographic affiliations of the studied populations. The present study, however, suggests relative homogeneity of the Meghalaya populations with a relatively low coefficient of gene differentiation (G_{ST}) (0.026 ± 0.002), which is lower than the value observed by Reddy et al. (2001c) for a substructured single caste population of Andhra Pradesh, based on 13 STR loci, some of which are common to the AmpFLSTR loci. This is naturally reflected in the lack of clear differentiation and clustering pattern of the Meghalaya tribes, either on geographic proximity or on the historical or current affiliations of these tribes. Even the position of the linguistically different Garo is not portrayed as distinct from the Khasi.

What could be the reasons for the reduced microsatellite diversity among the Meghalaya populations that results in the unclear pattern of population configurations? The following two plausible explanations can be advanced: (1) extensive admixture, both historical and current, among the Meghalaya tribes, including the Garo, which is perhaps perpetrated by the system of matrilineal descent and matrilocal residence; and (2) a relatively short history of separation of these tribes, which probably had a common and recent origin with negligible founder effects and subsequent genetic drift.

The evolutionary history of the process of substructuring of the Indian caste and tribal populations is expected to be relatively short. Therefore mutation cannot be considered the force behind the genetic differentiation of such populations. Most of the accrued genetic variation can at best be explained in terms of a founder effect in the process of formation of the subcastes and tribes and subsequent genetic drift over the generations. However, the history of substructuring of the castes is expected to be shorter than that of the autochthonous tribal populations. Therefore, if the barriers of subtribe endogamy are as rigid as those among the subcastes, then the tribes of Meghalaya show much greater differentiation and clearer formation of clusters than was apparent in the present study. This is especially pertinent because, as mentioned earlier, the hypervariable microsatellites bear the signatures of even the subcaste endogamy (Reddy et al. 2001c). This and the fact that even the Tibeto-Burman Garo have genetically integrated into the Khasi population rather than clustering with other Tibeto-Burman groups, including its own counterpart from north Bengal, prompt us to surmise that the short history of separation of the Meghalaya tribes is the real reason for the observed greater homogeneity. It is imperative that the answers be sought from the population structure of these tribes and the nature of marital interactions among them.

In the Meghalaya situation the subtribes of the Khasi are basically the regional and/or dialectal groups whose boundaries of endogamy do not seem to be absolutely impermeable. There are no rigid customary laws that prescribe marriages strictly within the notional boundaries of the subtribes. This is contrary to what has traditionally been the norm in the patrilineal tribal populations of at least mainland India. Implicit in the practice of matriarchy among these tribes is

the greater freedom and power of women in the household. Our personal knowledge and interaction with these populations suggest that matriarchy must have played a significant role in absorbing male spouses from outside the subtribe. In fact, the detailed demographic data generated from the Lyngngam, one of the subtribes of Meghalaya, indicate that there has been a prolonged and widespread inflow of genes from the neighboring Khasi as well as from the Garo tribes (Langstieh and Reddy 1999; Langstieh et al. 2003). There has also been consistent increase in this inflow during the last couple of decades with improvement in communications. Therefore we surmise that in this matriarchal society the influence of female preference over the selection of mates and inheritance of land and property revolving on matrilocal residence might have played a role in promoting genetic homogenization of the Meghalaya tribes. Therefore the reliability of phylogeny inferred from the presence of admixture could be profoundly compromised (Cavalli-Sforza et al. 1994; Perez-Lezaun et al. 1997; Chu et al. 1998), and this is probably reflected in the unintelligible pattern of population relationships in Meghalaya. Nevertheless, Meghalaya tribes as a whole maintain genetic identity compared with the populations from outside Meghalaya, as is evident from the compact cluster formed by them (Figure 6). This also suggests a relatively recent origin of these people from a common stock so that the differentiation is so low that it is overwhelmed by the degree of continuous gene flow.

This high degree of genetic admixture is also reflected in the uniformly high heterozygosity ($\cong$0.83) observed in almost all Meghalaya tribes, except the War Khasi, whose distribution in inaccessible and isolated areas promotes genetic isolation from the rest. In accordance with this, it is only the War Khasi that appear as a distinct outlier, albeit below the regression line and far removed from the centroid, suggesting the influence of genetic drift (Figure 4). This pattern of extreme admixture may also account for the observed departures from Hardy–Weinberg equilibrium. The structure analysis further suggests that only the War Khasi and the Pnar show any semblance of their genetic identity, and the membership coefficients of different populations of Meghalaya, including the Garo, indicate that there has been a considerable degree of admixture among them. In conclusion, the populations of Meghalaya are characterized by a relatively homogeneous genetic structure that is probably sustained by continuous gene flow across the tribal boundaries, and the matrilineal system prevalent among them might have played a role in promoting this gene flow.

Although the neighbor-joining tree (Figure 6) based on 9 STR loci among the 47 populations suggests closer affinity of the Meghalaya populations to the other Asian populations from East and Southeast Asia, including the Austro-Asiatic populations from Vietnam, Thailand, and Kaumpuchea, the final conclusions should await our present efforts in screening these populations for uniparently inherited mtDNA and Y-chromosome markers, which should provide better insights into the origin and genetic structure of these important tribal populations from northeastern India.

Acknowledgments This research is part of a project of the Indian Statistical Institute, Kolkata, which is being carried out in collaboration with the Centre for Cellular and Molecular Biology (CCMB), Hyderabad. We thank the directors of both institutes for logistic support. We thank A. Govardhan Reddy of CCMB for his technical inputs during the laboratory analysis.

Received 30 April 2003; revision received 1 December 2003.

Literature Cited

AmpFLSTR Profiler Plus PCR Amplification Kit Users Manual. 1998. Foster City, CA: PE Applied Biosystems.

Basu, A., N. Mukherjee, S. Roy et al. 2003. Ethnic India: A genomic view, with special reference to peopling and structure. *Genome Res*. 13:2,277–2,290.

Borys, S., A. Eisenberg, G. Carmody et al. 1999a. Allele frequencies for 9 STR loci in African-American, Chinese, Vietnamese, and Bangladesh populations. *J. Forensic Sci*. 44:1,316–1,318.

Borys, S., S. Iwaamoto, J. Miyakoshi et al. 1999b. Allele frequency distribution for 9 STR loci in the Japanese population. *J. Forensic Sci*. 44:1,319.

Bowcock, A.M., A. Ruiz-Linares, J. Tomfohrde et al. 1994. High-resolution evolutionary trees with polymorphic microsatellites. *Nature* 368:455–457.

Cavalli-Sforza, L.L., P. Menozzi, and A. Piazza. 1994. *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press.

Chattopadhyay, P., R. Dutta, and V.K. Kashyap. 2001. Allele frequency data at nine fluorescent labeled STR loci in four tribal population groups of India. *J. Forensic Sci*. 45:184–188.

Chu, J.Y., W. Huang, S.Q. Kuang et al. 1998. Genetic relationship of populations in China. *Proc. Natl. Acad. Sci. USA* 95:11,763–11,768.

Deka, R., M.D. Shriver, L. Jin et al. 1995a. Population genetics of dinucleotide (dC-dA)n.(dG-dT)n. polymorphisms in world populations. *Am. J. Hum. Genet*. 56:461–474.

Deka, R., M.D. Shriver, L.M. Yu et al. 1995b. Intra- and interpopulation diversity at short tandem repeat loci in diverse populations of the world. *Electrophoresis* 16:1,659–1,664.

Dutta, R., P. Chattopadhyay, and V.K. Kashyap. 2001. STR data for the AmpF*l*STR profiler plus loci among four predominant populations of eastern India. *J. Forensic Sci*. 45:1,353–1,357.

Dutta, R., B.M. Reddy, P. Chattopadhyay et al. 2002. Patterns of genetic diversity at the nine forensically approved STR loci in the Indian populations. *Hum. Biol*. 74:33–49.

Eller, E. 1999. Population substructure and isolation by distance in three continental regions. *Am. J. Phys. Anthropol*. 108:147–159.

Excoffier, L., P.E. Smouse, and J.M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131:479–491.

Falush, D., M. Stephens, and J.K. Pritchard. 2003. Inferences of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164:1,567–1,587.

Fung, W.K., J. Ye, L. Hu et al. 2001. Allele frequencies for nine STR loci in Beijing Chinese. *Forensic Sci. Int*. 121:207–209.

Guo, S., and E. Thompson. 1992. Performing exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* 48:361–372.

Harpending, H.C., and R. Ward. 1982. Chemical systematics and human evolution. In *Biochemical Aspects of Evolutionary Biology,* M. Nitecki, ed. Chicago: University of Chicago Press, 213–256.

Hussain, Z. 1991. Who are the prehistoric dwellers of the Meghalaya Plateau? In *Archaeology of Northeast India,* J.P. Singh and G. Sengupta, eds. New Delhi: Vikas Publishing, 74–85.

Jorde, L.B., A.R. Rodgers, M. Bamshad et al. 1997. Microsatellite diversity and the demographic history of modern humans. *Proc. Natl. Acad. Sci. USA* 94:3,100–3,103.

Langstieh, B.T., and B.M. Reddy. 1999. The origin and ethnic position of Lyngngam among the tribes of Meghalaya: An exploratory study. *J. Ind. Anthropol. Soc*. 34:265–275.

Langstieh, B.T., B.M. Reddy, K. Thangaraj et al. 2003. Disputed origin of Lyngngam among the tribes of Meghalaya: Can molecular genetic markers resolve? Unpublished manuscript.

Maniatis, T., J. Sambrook, and E.F. Fritsch. 1989. *Molecular Cloning: A Laboratory Manual,* 2nd ed. Cold Spring Harbor, NY: Laboratory Press.

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

Nei, M., F. Tajima, and Y. Tateno. 1983. Accuracy of estimated phylogenetic trees from molecular data. *J. Mol. Evol*. 19:153–170.

Nei, M., and N. Takezaki. 1996. The root of the phylogenetic tree of human populations. *Mol. Biol. Evol*. 13:170–177.

Parra, E., N. Saha, A.G. Soemantri et al. 1999. Genetic variation at 9 autosomal microsatellite loci in Asian and Pacific populations. *Hum. Biol*. 71:757–779.

Perez-Lezaun, A., F. Calafell, E. Mateu et al. 1997. Microsatellite variation and differentiation of modern humans. *Hum. Genet*. 99:1–7.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.

Reddy, B.M., R. Dutta, B. Langstieh et al. 2001a. Diversity at three tetrameric STR loci in a substructured Golla caste population of southern Andhra Pradesh in comparison to other Indian populations. *Int. J. Hum. Genet*. 1:1–9.

Reddy, B.M., G. Sun, and R. Dutta. 2001b. STR data for the AmpF*l*STR Profiler Plus loci among Golla population of southern Andhra Pradesh, India. *J. Forensic Sci*. 46:734–735.

Reddy, B.M., G. Sun, J.R. Luis et al. 2001c. Genomic diversity at thirteen short tandem repeat loci in a substructured caste population, Golla, of southern Andhra Pradesh, India. *Hum. Biol*. 73:175–190.

Relethford, J.H. 1993. *Mantel for Windows, Ver. 2.0*. Available from the author (relethjh@oneonta.edu).

Reynolds, J., B.S. Weir, and C.C. Cockerham. 1983. Estimation for coancestry coefficient: Basis for a short-term genetic distance. *Genetics* 105:767–779.

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

Schneider, S., D. Rosslie, and L. Excoffier. 1997. *Arlequin, ver 2.000: A Software for Population Genetics Data Analysis*. Geneva: Genetics and Biometry Laboratory, University of Geneva. Available at http://anthropologie.unige.ch/arlequin

Sharma, T.C. 1966. Researches on the prehistoric archeology of Assam. *J. Assam. Sci. Soc*. 9:1–11.

Sharma, T.C. 1980. Prehistoric archeology in northeastern India: A review in progress. In *Eastern Himalayas,* T.C. Sharma and D.N. Majumdar, eds. New Delhi: Cosmo Publications, 102–107.

Shriver, M.D., M.W. Smith, L. Jin et al. 1995. Ethnic-affiliation estimation by use of population-specific DNA markers. *Am. J. Hum. Genet*. 60:957–964.

Stoneking, M., J.J. Fontius, S.L. Clifford et al. 1997. *Alu* insertion polymorphisms and human evolution: Evidence for a larger population size in Africa. *Genome Res*. 7:1,061–1,071.

Sun, G., S.T. McGarvey, R. Bayoumi et al. 2003. Global genetic variation at nine short tandem repeat loci and implications on forensic genetics. *Eur. J. Hum. Genet*. 11:39–49.

Takezaki, N., and M. Nei. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144:389–399.

Appendix 1. Allele Frequencies for the Nine Autosomal Loci Among the Nine Meghalayan Populations[a]

Allele	*Lyngngam* ($N = 156$)	*Garo* ($N = 128$)	*Nongtrai* ($N = 90$)	*Maram* ($N = 96$)	*Khynriam* ($N = 146$)	*Pnar* ($N = 100$)	*Bhoi* ($N = 90$)	*War Khasi* ($N = 80$)	*War Jaintia* ($N = 46$)
D3S1358									
12	–	–	0.0111	–	–	–	–	–	–
13	–	–	0.0111	0.0104	–	–	–	–	–
14	0.0256	0.0313	0.0333	0.0104	0.0068	0.0200	0.0111	0.0125	0.0217
15	0.2821	0.2656	0.2000	0.2604	**0.3356**	**0.2900**	0.2778	**0.3125**	**0.3478**
15.2	–	0.0078	0.0111	–	0.0068	0.0300	–	–	–
16	**0.4038**	**0.4453**	0.2778	**0.3646**	0.2877	**0.2900**	**0.3556**	**0.3125**	0.3261
17	0.2436	0.1094	0.3667	0.2292	0.1781	0.2000	0.2000	0.2125	0.1087
18	0.0385	0.1328	0.0889	0.1042	0.1575	0.1600	0.1333	0.1375	0.1957
19	0.0064	0.0078	–	0.0104	0.0274	0.0100	0.0111	0.0125	–
20	–	–	–	0.0104	–	–	0.0111	–	–
D8S1179									
8	0.0385	0.0313	–	–	–	0.0100	–	0.0500	–
9	0.0128	0.0468	0.0556	0.0417	–	0.0200	–	0.0125	–
10	0.0577	0.0547	0.0556	0.1250	0.1644	0.1100	0.0889	0.1375	0.1522
11	0.1410	0.0547	0.1333	0.1250	0.0959	0.1200	0.0556	0.1125	0.1522
12	0.1410	0.1641	0.1111	0.0833	0.1370	0.0900	0.0444	**0.3000**	**0.2609**
13	0.0962	0.1406	0.2000	0.1667	**0.2123**	0.1700	0.1778	0.0625	0.1522
14	**0.2372**	**0.2188**	**0.2222**	0.1354	0.1096	**0.2200**	**0.2889**	0.0625	0.0435
15	0.1538	0.1406	0.1000	**0.2083**	0.1712	0.1300	0.2000	0.1250	0.1087
16	0.0833	0.1250	0.0778	0.0521	0.1027	0.1100	0.1222	0.1250	0.1087
17	0.0385	0.0156	0.0333	0.0521	0.0068	0.0200	0.0111	0.0125	0.0217
18	–	0.0078	0.0111	0.0104	–	–	0.0111	–	–
D5S818									
7	0.0064	0.0156	0.0444	–	–	0.0100	0.0111	0.0250	–
8	–	–	–	–	–	0.0100	–	–	–
9	0.0769	0.0547	0.0889	0.0313	0.0411	0.0300	0.0556	0.0750	0.0870
10	0.1795	0.2500	0.1667	0.1979	0.2603	0.2700	0.2222	**0.3000**	**0.3043**
11	**0.3205**	**0.3203**	**0.3889**	**0.4063**	**0.2671**	**0.4100**	**0.3667**	0.2375	0.2609
12	0.1731	0.1797	0.1667	0.1875	0.2603	0.1300	0.1333	0.2750	0.2609
13	0.2308	0.1094	0.1444	0.1667	0.1644	0.1400	0.1667	0.0625	0.0869
14	0.0128	0.0547	–	0.0104	0.0068	–	0.0222	0.0250	–
15	–	–	–	–	–	–	0.0222	–	–
16	–	0.0156	–	–	–	–	–	–	–
VWA									
11	–	–	–	–	–	0.0100	–	0.0250	–
12	–	–	–	0.0208	–	–	0.0111	0.0125	–
13	–	0.0078	–	–	–	–	–	0.0125	–
14	0.1731	0.1406	0.0667	0.1458	0.1233	0.0400	0.0667	–	0.0652
15	0.0449	0.0156	0.0222	0.0625	0.0274	0.0500	0.0111	–	0.0217
15.2	0.0577	0.0234	0.0111	–	–	–	0.0111	0.0125	–
16	0.1218	0.1797	**0.2556**	0.1458	**0.2740**	**0.3000**	**0.2444**	**0.3625**	**0.2609**
17	**0.2692**	**0.3203**	**0.2556**	**0.2292**	0.2397	0.2500	0.2333	0.1750	**0.2609**
18	0.1731	0.1953	0.2556	0.1875	0.2329	0.2300	0.2222	0.2000	0.2174
19	0.1410	0.0781	0.1333	0.1250	0.0753	0.1000	0.1778	0.1375	0.1739
20	0.0192	0.0078	–	0.0521	0.0205	0.0100	0.0222	0.0625	–

Appendix 1. Continued.

Allele	*Lyngngam* ($N = 156$)	*Garo* ($N = 128$)	*Nongtrai* ($N = 90$)	*Maram* ($N = 96$)	*Khynriam* ($N = 146$)	*Pnar* ($N = 100$)	*Bhoi* ($N = 90$)	*War Khasi* ($N = 80$)	*War Jaintia* ($N = 46$)
21	–	0.0313	–	0.0208	0.0068	–	–	–	–
22	–	–	–	0.0104	–	0.0100	–	–	–
D21S11									
24.2	0.0064	–	0.0111	0.0104	–	–	–	–	–
25	0.0064	0.0078	–	–	–	–	0.0111	–	–
25.3	–	–	–	0.0104	–	–	–	–	–
26	–	0.0078	–	–	–	–	–	–	–
27	–	0.0313	–	0.0104	–	–	0.0333	–	–
28	0.0385	0.0469	–	0.0938	0.0411	0.1400	0.0333	0.0125	0.0217
28.2	0.0064	0.0156	–	–	0.0479	–	–	0.0125	–
28.3	0.0833	0.0234	0.0111	0.0417	0.0068	0.1200	0.0333	–	0.0217
29	**0.1282**	**0.1719**	0.0444	**0.1875**	0.1781	0.0600	0.2000	0.1625	0.1087
29.1	0.0769	0.0313	0.1778	0.0104	0.0274	0.0100	0.0667	0.0625	0.0870
29.2	–	0.0156	–	–	0.0205	–	0.0111	0.0125	0.0435
29.3	0.0705	0.0156	–	–	0.0068	0.1300	–	–	–
30	0.0769	0.0781	0.1222	0.0104	**0.2397**	0.0600	0.0778	**0.3125**	0.1739
30.1	0.0385	0.1406	0.1111	**0.1875**	0.0685	0.0100	0.0889	0.0625	0.0652
30.2	0.0128	0.0078	–	–	0.0137	–	0.0667	0.0125	0.0217
30.3	0.0128	0.0938	–	0.0521	0.0137	–	0.0333	0.0125	–
31	0.0577	–	0.0222	0.0313	0.0137	0.0600	0.0333	0.0125	0.0217
31.1	0.0192	0.0234	0.0889	0.0208	0.0342	0.0400	0.0111	–	–
31.2	0.0385	–	0.0111	–	0.0137	0.0300	–	0.0250	0.0217
32	0.0513	0.0781	0.0222	0.0104	0.0411	–	0.0333	0.0500	0.0435
32.1	0.0192	0.0078	0.0111	0.0104	0.0137	0.0100	0.0111	–	–
32.2	0.1026	0.0313	0.0556	0.0313	0.0548	**0.1900**	0.0556	0.0750	**0.2174**
33	0.1218	0.1172	**0.1889**	0.1458	0.1438	0.0300	0.1000	0.1000	0.1087
33.1	0.0064	0.0078	0.0111	0.0104	–	–	–	0.0375	–
33.2	0.0128	–	0.0111	–	0.0068	0.0700	0.0222	–	–
33.3	0.0128	–	0.0333	0.0104	0.0068	–	0.0333	0.0125	0.0435
34	–	–	0.0222	–	–	0.0100	0.0111	–	–
34.1	–	–	0.0222	–	–	–	–	–	–
35	–	0.0234	0.0111	0.0625	–	–	–	–	–
35.2	–	0.0156	0.0111	–	–	0.0100	–	–	–
36	–	–	–	0.0208	0.0068	0.0100	0.0333	0.0250	–
36.2	–	–	–	0.0104	–	–	–	–	–
38	–	0.0078	–	0.0208	–	0.0100	–	–	–
D13S317									
8	**0.2756**	0.0781	0.1222	**0.2708**	0.2055	**0.2200**	**0.2222**	**0.3000**	0.1739
9	0.1603	0.1953	0.0778	0.1042	0.1027	0.1100	0.1889	0.1625	0.1304
10	0.0962	0.1719	0.1556	0.1042	0.0822	0.1700	0.1333	0.2250	0.1087
11	0.0577	0.1250	**0.2000**	0.1875	0.1233	0.1400	0.0778	0.2750	0.0435
12	0.1795	**0.2188**	0.1889	0.1875	**0.2534**	0.1800	**0.2222**	0.0250	**0.3478**
13	0.1731	0.0781	0.1778	–	0.1507	0.1500	0.1111	0.0125	0.1087
14	0.0513	0.0938	0.0556	0.0521	0.0753	0.0200	0.0333	–	0.0870
15	0.0064	0.0391	0.0222	0.0938	0.0068	0.0100	0.0111	–	–
FGA									
16	–	–	–	–	–	–	–	0.0250	–

Allele	*Lyngngam* ($N = 156$)	*Garo* ($N = 128$)	*Nongtrai* ($N = 90$)	*Maram* ($N = 96$)	*Khynriam* ($N = 146$)	*Pnar* ($N = 100$)	*Bhoi* ($N = 90$)	*War Khasi* ($N = 80$)	*War Jaintia* ($N = 46$)
17	–	–	–	–	–	–	–	0.0250	–
18	0.0641	0.0078	0.0444	0.0104	0.0068	0.0200	0.0222	0.0125	0.0435
18.2	0.0064	0.0078	0.0333	–	0.0137	–	–	0.0125	0.0217
19	0.1090	0.0547	0.1444	0.0625	0.0479	0.0200	0.0333	0.0250	–
19.2	0.0256	0.0781	–	–	0.0411	–	0.0111	–	–
20	0.0513	0.0391	0.0667	0.0417	0.0205	0.1000	0.0444	–	0.0217
20.2	0.0192	0.0391	0.0222	0.0208	0.0548	0.0400	–	–	0.0217
21	0.1154	0.0625	0.0222	0.1250	0.0890	0.0300	0.0444	0.0875	0.0870
21.2	0.0449	0.0703	0.0222	0.0208	**0.1301**	**0.1500**	0.0667	–	**0.1522**
22	**0.1859**	0.0625	0.0889	0.1458	0.0137	0.0100	**0.1444**	0.1250	0.0 217
22.2	0.0192	0.0625	0.0111	0.0208	0.0890	0.1400	0.0667	0.0375	0.0870
22.3	–	0.0234	0.0111	–	0.0137	–	0.0111	–	–
23	0.1026	0.0078	0.0667	0.0104	0.0685	0.0100	0.0889	**0.3500**	0.1087
23.2	0.0577	0.0156	0.0111	**0.2500**	0.0685	0.0800	0.1000	0.0375	0.0652
24.0	0.0449	0.1172	**0.1667**	0.0208	0.0685	0.0400	0.1000	0.0250	0.0652
24.2	0.0128	0.0156	0.0778	0.0729	0.1164	0.1200	0.0889	0.0375	0.0870
25	0.0641	0.0703	0.0778	0.0104	0.0616	0.0900	0.0333	0.0125	**0.1522**
25.2	0.0192	–	0.0111	0.0521	0.0274	0.0700	0.1000	0.0125	–
26	0.0385	**0.1484**	0.0333	0.0104	0.0137	0.0600	0.0333	0.0625	–
27	0.0192	0.0391	0.0778	0.0521	0.0205	–	0.0111	0.1000	0.0435
27.2	–	0.0078	0.0111	0.0104	0.0342	–	–	0.0125	–
28	–	0.0234	–	0.0521	–	0.0200	–	–	0.0217
28.2	–	0.0078	–	0.0104	–	–	–	–	–
29	–	0.0078	–	–	–	–	–	–	–
30	–	0.0313	–	–	–	–	–	–	–
D7S820									
6	0.0064	0.0156	–	0.0417	–	–	–	–	–
7	–	–	0.0333	–	0.0068	–	–	–	–
8	0.1795	0.0547	**0.3222**	0.1563	0.1712	0.2000	0.1889	0.0125	0.0652
9	0.0897	0.2031	0.0889	0.0417	0.0822	0.0800	0.1556	**0.2625**	0.1739
10	0.2115	0.1563	0.0778	0.2396	0.1438	**0.2600**	0.1000	0.1375	0.1304
11	0.2244	0.2109	0.2778	0.1875	**0.2740**	0.1800	0.2000	0.2250	**0.2391**
12	**0.2500**	**0.2813**	0.0889	**0.2917**	0.2329	0.2400	**0.2333**	0.1625	0.2174
13	0.0321	0.0391	0.0667	0.0313	0.0685	–	0.0778	0.2000	0.1522
14	0.0064	0.0234	0.0222	0.0104	0.0068	0.0300	0.0111	–	–
15	–	0.0156	0.0222	–	0.0137	0.0100	0.0333	–	0.0217
D18S51									
10	–	–	–	–	0.0137	–	–	0.0500	–
11	0.0064	–	–	0.0104	–	–	0.0222	0.0125	–
12	0.0192	0.0156	0.0333	0.0313	0.0068	0.0400	–	0.0500	0.0652
13	0.0962	0.0625	0.0667	0.1250	0.0753	0.1500	0.0889	0.1000	0.1087
13.2	0.0577	0.0781	0.0111	0.0104	0.1575	–	0.0222	0.0125	–
14	0.1346	0.1172	0.1556	0.1042	0.0137	0.2000	0.1444	0.0875	0.0652
14.2	0.0705	0.0234	0.0222	0.0104	0.1096	–	0.0111	–	–
15	**0.2500**	**0.2500**	**0.4000**	**0.2188**	–	**0.2700**	**0.2667**	**0.5000**	**0.3260**
15.2	0.0385	0.0859	–	–	**0.1849**	0.0200	0.0333	–	0.0435
16	0.0962	0.0781	0.1556	0.1250	0.0274	0.0800	0.1222	0.1125	0.0217

Appendix 1. Continued.

Allele	*Lyngngam* ($N = 156$)	*Garo* ($N = 128$)	*Nongtrai* ($N = 90$)	*Maram* ($N = 96$)	*Khynriam* ($N = 146$)	*Pnar* ($N = 100$)	*Bhoi* ($N = 90$)	*War Khasi* ($N = 80$)	*War Jaintia* ($N = 46$)
17	0.1538	0.0703	–	0.0938	0.1712	0.0800	0.0778	–	0.0652
17.2	–	0.0234	0.0111	0.0208	0.0342	0.0100	–	–	–
18	0.0385	0.0156	0.0556	0.0417	0.0137	0.0400	0.0111	0.0375	0.0652
19	0.0128	0.0625	0.0111	0.0833	0.0342	0.0200	0.0444	0.0375	0.1739
19.2	–	0.0625	–	0.0104	0.0479	0.0500	–	–	–
20	0.0064	0.0313	0.0111	0.0313	–	–	0.1000	–	0.0217
21	0.0192	0.0078	0.0222	–	0.0616	0.0300	0.0111	–	0.0217
22	–	0.0156	0.0333	0.0208	0.0342	–	–	–	0.0217
23	–	–	–	0.0104	0.0068	0.0100	–	–	–
24	–	–	0.0111	0.0208	–	–	0.0222	–	–
25	–	–	–	0.0208	–	–	0.0222	–	–
26	–	–	–	0.0104	0.0068	–	–	–	–

a. The values in boldface are the modal frequencies.