Genetic variation at twentythree microsatellite loci in sixteen human populations

RANJAN DEKA[1*], MARK D. SHRIVER[2], LING MEI YU[3], ELISA MUELLER HEIDREICH[3], LI JIN[4], YIXI ZHONG[4], STEPHEN T. McGARVEY[5], SHYAM SWARUP AGARWAL[6], CLAREANN H. BUNKER[7], TETSURO MIKI[8], JOACHIM HUNDRIESER[9], SHIH-JIUN YIN[10], SALMO RASKIN[11], RAMIRO BARRANTES[12], ROBERT E. FERRELL[3] and RANAJIT CHAKRABORTY[4]

[1]*Department of Environmental Health, University of Cincinnati, Cincinnati, Ohio, USA;* [2]*Department of Anthropology, Pennsylvania State University, University Park, Pennsylvania, USA;* [3]*Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA;* [4]*Human Genetics Center, University of Texas Health Science Center, Houston, Texas, USA;* [5]*Department of Medicine and International Health Institute, Brown University, Providence, Rhode Island, USA;* [6]*Sanjay Gandhi Post-Graduate Institute of Medical Sciences, Lucknow, India;* [7]*Department of Epidemiology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA;* [8]*Department of Geriatric Medicine, Ehime University, Ehime, Japan;* [9]*Klinik für Abdominal und Transplantationschirurgie, Medizinische Hochschule, Hannover, Germany;* [10]*Department of Biochemistry, National Defense Medical Center, Taipei, Taiwan;* [11]*Universidade Federal do Parana, Curitiba, Brazil;* [12]*Instituto de Investigaciones en Salud, Universidad de Costa Rica, San Jose, Costa Rica*

Abstract

We have analysed genetic variation at 23 microsatellite loci in a global sample of 16 ethnically and geographically diverse human populations. On the basis of their ancestral heritage and geographic locations, the studied populations can be divided into five major groups, viz. African, Caucasian, Asian Mongoloid, American Indian and Pacific Islander. With respect to the distribution of alleles at the 23 loci, large variability exists among the examined populations. However, with the exception of the American Indians and the Pacific Islanders, populations within a continental group show a greater degree of similarity. Phylogenetic analyses based on allele frequencies at the examined loci show that the first split of the present-day human populations had occurred between the Africans and all of the non-African populations, lending support to an African origin of modern human populations. Gene diversity analyses show that the coefficient of gene diversity estimated from the 23 loci is, in general, larger for populations that have remained isolated and probably of smaller effective sizes, such as the American Indians and the Pacific Islanders. These analyses also demonstrate that the component of total gene diversity, which is attributed to variation between groups of populations, is significantly larger than that among populations within each group. The empirical data presented in this work and their analyses reaffirm that evolutionary histories and the extent of genetic variation among human populations can be studied using microsatellite loci.

[Deka R., Shriver M. D., Yu L. M., Heidreich E. M., Jin. L., Zhong Y. *et al.* 1999 Genetic variation at twentythree microsatellite loci in sixteen human populations. *J. Genet.* **78**, 99–121]

Introduction

In recent years a large volume of genetic data has been generated on a global scale in efforts to understand the evolutionary histories and relationships of contemporary human populations. This has primarily been facilitated by the advent of PCR technology and molecular characterization of a large number of polymorphic loci in the human genome. Among these loci, the 'microsatellites' or 'short tandem repeats' (STR), characterized by length variation in tandem arrays of simple repeat sequences ranging from two to five nucleotides, are the most abundant in the genome.

* For correspondence: Department of Environmental Health, University of Cincinnati, P.O. Box 670056, 3223 Eden Avenue, Cincinnati, OH 45267-0056, USA. E-mail: ranjan.deka@uc.edu.

Keywords. microsatellite loci; genetic variation; gene diversity; human populations.

Their associated hypervariability has been the key in the success of the human genome project. Although these loci are subject to recombination and high mutation rates associated with a convergent nature of mutational process (Shriver *et al.* 1993; Weber and Wong 1993; Di Rienzo *et al.* 1994), recent studies have shown that microsatellites are also powerful tools in inferring evolutionary relationships and demographic histories of human populations (Bowcock *et al.* 1994; Deka *et al.* 1995a, 1995b, 1998; Jorde *et al.* 1995; Pérez-Lezaun *et al.* 1997; Shriver *et al.* 1997; Calafell *et al.* 1998; Kimmel *et al.*1998) and in the determination of parentage and relatedness of individuals (Chakraborty and Jin 1993). In this context it should be noted that the stepwise nature of mutations at microsatellite loci is amenable to rigorous mathematical and simulation studies for assessing evolutionary histories of populations (Kimmel and Chakraborty 1996; Kimmel *et al.* 1996; Shriver *et al.* 1997). In the study reported here we have analysed genetic variation at 23 microsatellite loci in 16 diverse human populations to examine the key issues of phylogenetic relationships of modern human populations and the extent of genetic diversity among these extant populations. In appendix we have provided the allele frequency data at the examined loci, which could be used by other investigators.

Materials and methods

Population samples: Blood or DNA samples were collected from about 800 individuals representing 16 ethnically and geographically diverse human populations. On the basis of their ancestral heritage and geographical locations, these populations can be divided into five major groups: (i) Africans, represented in this study by two populations from West Africa, namely Nigerian and Benin, both belonging to the Hausa group of tribes from Nigeria, and a third population from Brazil, whom we have named Brazilian Black; (ii) Caucasians, represented by a German sample from northern Germany, unrelated parents of the CEPH (Centre d'Etude du Polymorphisme Humain) cohort, a sample of unrelated whites from Brazil, the Brazilian White, and a caste population from northern India, the Brahmin; (iii) Asian Mongoloids, represented by a Chinese sample of Han origin, a Japanese population from Osaka, and a tribal population from northeast India, the Kachari; (iv) American Indians, consisting of the Dogrib Indian from the Northwest Territories of Canada, the Pehuenche Indian from southern Chile, and the Cabecar from Costa Rica; (v) The Pacific Islanders, represented by two groups of Samoans, American and Western, drawn from villages distributed throughout

Table 1. Summary of the 23 microsatellite loci.

Locus	Chromosomal location	Repeat motif	CEPH reference genotype[a] 133101	133102
Dinucleotide repeats				
D13S71	13q32-q33	CA	75/75 (17/17)	75/75 (17/17)
D13S118	13q14	CA	190/194 (17/19)	190/190 (17/17)
D13S121	13q31	CA	168/170 (20/21)	162/170 (17/21)
D13S122	13q31-q32	CA	87/97 (10/15)	87/107 (10/20)
D13S124	13q21	CA	185/191 (15/18)	185/185 (15/15)
D13S193	13q31-q32	CA	147/147 (22/22)	145/147 (21/22)
D13S197	13q31-q32	CA(GC)CA	97/97[b]	126/128
FLT1	13q12	CA	170/182 (16/22)	168/168 (15/15)
Trinucleotide repeats				
PLA2A1	12q23-qter	AAT	130/133 (14/15)	121/121 (11/11)
D20S473	20p13	ATA	181/181 (12/12)	181/181 (12/12)
Tetranucleotide repeats				
TH01	11p15.5	AATG	184/199 (6/9.3)	184/199 (6/9.3)
CSF1R	5q33.3-q34	AGAT	311/323 (11/14)	315/315 (12/12)
F13A1	6p24-p25	AAAG	283/295 (4/7)	287/295 (5/7)
CYP19	15q21.1	AAAT	173/197 (5/11)	173/193 (5/10)
LPL	8p22	AAAT	123/131 (10/12)	123/127 (10/11)
D20S604	20p11.2-12	GATA	135/147 (13/16)	139/139 (14/14)
D20S481	20q11.2-12	GATA	237/241 (15/16)	233/245 (14/17)
D21S1435	21q21	GATA	171/179 (10/12)	175/175 (11/11)
D21S1446	21q22	GATA	209/223[c]	223/223
Disease-associated trinucleotide repeats				
DM	19q13.3	CTG	78/111 (5/16)	78/78 (5/5)
SCA1	6p22-23	CAG(CAT)CAG	217/217 (30/30)	223/223 (32/32)
DRPLA	12pter-12	CAG	143/146 (15/16)	143/146 (15/16)
HD	4p16.3	CAG	122/140 (18/24)	113/137 (15/23)

[a] Genotype of CEPH individuals 133101 and 133102 are given as reference markers in terms of the fragment sizes in base pairs and the corresponding repeat numbers in parentheses; [b]fragment sizes at D13S197 do not always correspond to a two-base-pair increment and therefore the corresponding repeat sizes could not be deduced; [c]fragment sizes at D21S1446 also do not correspond to four-base-pair increment and therefore repeat numbers could not be deduced. See text for details.

American and Western Samoa, and the New Guinea Highlander from the northern fringes of Papua New Guinea.

Laboratory analysis: A summary of the microsatellite loci is presented in table 1. Of the 23 loci, eight are dinucleotide repeats, six are trinucleotides, and nine are tetranucleotide repeats. Four (DM, SCA1, DRPLA and HD) of the trinucleotide repeats are associated with human diseases. The details of the PCR amplification and analysis of the dinucleotide repeats are given in Deka *et al.* (1995a). Eleven of the trinucleotide and tetranucleotide repeat loci were analysed in four standard multiplex PCR reactions: (i) CSF1R, TH01 and PLA2A; (ii) F13A1, CYP19 and LPL; (iii) D21S1446 and D21S1435; and (iv) D20S481, D20S473 and D20S604. The four disease-associated trinucleotide repeats, viz. DM, SCA1, DRPLA and HD, were amplified individually. Twentyfive to fifty nanograms of DNA was amplified in a total volume of 25 µl reaction mixture containing standard PCR buffer (10 mM Tris-HCl, pH 8.3; 50 mm KCl; 1.5 mM $MgCl_2$), 200 µM each dNPT, one unit of Taq DNA polymerase, unlabelled (1 µM) and labelled primers. The forward primers were end-labelled using [γ-^{33}P] ATP and polynucleotide kinase T4. The amplified products were separated on 6% denaturing polyacrylamide gels. Following electrophoresis, the gels were dried and allelic fragments were visualized by autoradiography. An M13 sequence ladder was run on each gel as a size standard. Finally, the repeat sizes were determined by sequencing a subset of alleles (CSF1R, TH01, PLA2A, F13A1, CYP19, LPL, D20S473) and from published sequences (DM, SCA1, DRPLA, HD) or inferring from Genbank sequences (FLT1, D13S118, D13S121, D13S71, D13S122, D13S193, D13S124, D21S1435, D20S481 and D20S604). Alleles at these loci have been designated by the size (in base pairs) of their PCR product and their corresponding repeat numbers (see table 1 and appendix). Fragment sizes at the D13S197 (a dinucleotide repeat) and D21S1446 (a tetranucleotide repeat) loci do not always correspond to two-bp and four-bp increments, respectively. Therefore, alleles at these two loci could not be assigned repeat numbers.

Data analysis: As all the examined loci are autosomal and detect codominant alleles, allele frequencies were calculated by the gene counting method (Li 1976). The apportionment of genetic variability in the total data set was done by the method of Chakraborty (1980), following the concept of gene diversity of Nei (1973). Genetic distances were estimated by using the modified Cavalli-Sforza distance measure (D_A, Nei *et al.* 1983) and the stepwise-weighted D_{SW} measure (Shriver *et al.* 1995). D_{SW} is essentially a modification of Nei's minimum distance measure, which takes into account both allele frequency differences as well as allele size differences. Neighbour-joining trees (Saitou and Nei 1987) were constructed on the basis of both these distance measures, and the degree of support of the branches was evaluated by a bootstrap analysis based on 1000 replications.

Results

Allelic distributions

Allele frequencies at the 23 microsatellite loci in the 16 studied populations are presented in the appendix. The number of chromosomes examined from each population is shown in the last row for each locus. For example, 220 chromosomes from the Nigerian population were analysed at the D13S71 locus. Although a comprehensive locus by locus discussion of the allelic distributions is not the objective of this report, a few salient features are noteworthy. The spectrum of allelic distributions across the populations is quite broad and significant interpopulation variation at these loci is often observed. The tetranucleotide repeat loci TH01 and CSF1R are two illustrative examples of contrasting patterns of such variations. At the TH01 locus, the allele frequency distributions are quite varied across populations. However, a distinctive feature emerges when observations are made with respect to the groupings of the major populations. In general, populations within a major group show a greater degree of similarity. For example, repeat 9 is the predominant allele in every Asian Mongoloid population with frequencies of 40–52%, while this allele is present at considerably lower frequencies in all of the African and the Caucasian populations with a range between 5 and 20%. On the other hand, repeat 9.3 is the predominant allele among the Caucasians (35–40%), with the exception of the Brahmin population from India (13%). This allele is present at much lower frequencies in all of the Asian Mongoloid and African populations (1–13%). In contrast, at the CSF1R locus allele frequencies are more uniform across populations irrespective of their geographic or ethnic ancestry. There are three major alleles (repeats 10, 11 and 12) in all populations.

Clustering of populations within a major group as noted above, however, is not evident among populations classified under the major groups of American Indians and Pacific Islanders. For instance, at the TH01 locus again, repeat 6 is the predominant allele (68%) among the Cabecar. In contrast, its frequency is 34% among the Pehuenche Indians, and among the Dogribs it is rare, with a frequency of just about 2%. The Pacific Island populations, the New Guinea Highlanders and the two Samoan groups show significantly different allele frequencies at all loci. It should also be noted that the New Guinea Highlanders and the Cabecars from Costa Rica are also characterized by presence of a few high-frequency alleles at several loci. For example, among the New Guinea Highlander the frequency of the predominant allele at five loci (PLA2A, CYP19, F13A1, D20S473 and HD) is $\geq$80%. Similarly, the Cabecars have such high-frequency ($\geq$80%) alleles at four loci (FLT1, D13S197, D20S473 and DM). In fact, they are

monomorphic at the FLT1 locus. In general, however, the FLT1 locus shows a reduced level of variation in most populations (see discussion section).

As noted above, four of the studied loci have been implicated in human diseases, viz. myotonic dystrophy (DM), spino-cerebellar ataxia type 1 (SCA1), dentatorubral pallidoluysian atrophy (DRPLA) and Huntington disease (HD). The trinucleotide repeats associated with these loci are normally polymorphic in human populations. They result in disease when the repeat length exceeds a certain threshold. In our sample of populations, we have not observed any allele in the disease-causing range. Therefore the distribution of alleles is reflective of normal polymorphism at these loci in human populations. In general, we have observed a higher level of genetic variation at these four loci than at the other microsatellites. The number of alleles across the non-disease associated loci (with the exception of D13S197, which has 30 different alleles) ranges between 9 and 16 (average 11.2 ± 0.64). At the four disease-associated loci, this range is 19 to 28 (average 23.0 ± 2.12). As discussed below, this observation is also apparent in the analysis of gene diversity.

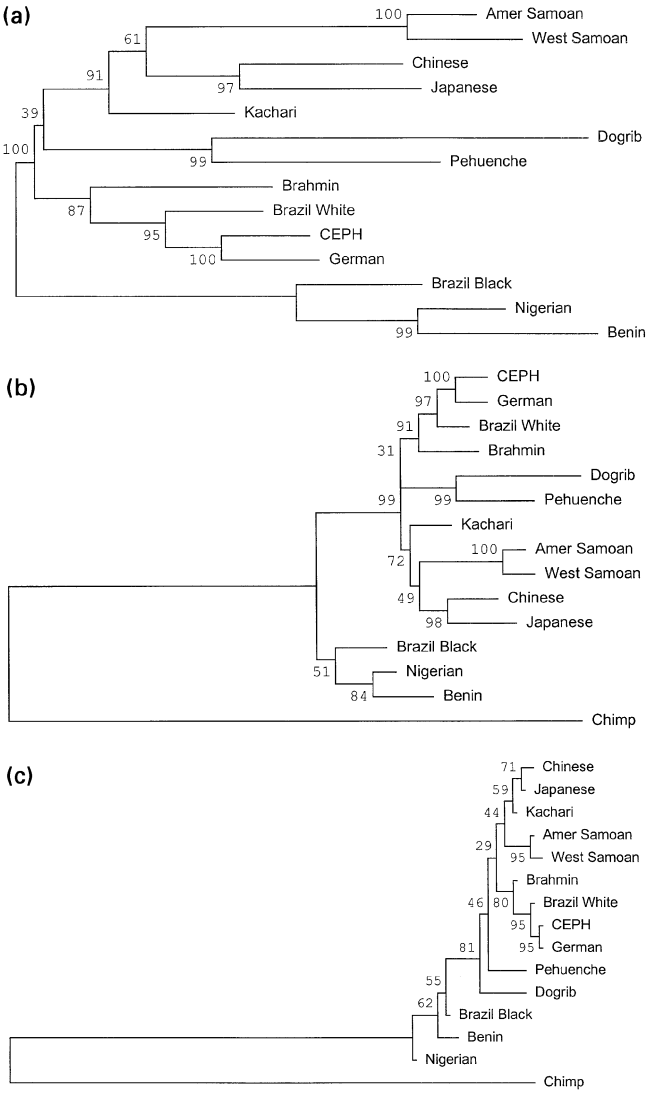

Figure 1. Phylogenetic trees of 14 human populations based on 23 microsatellite loci: (a) neighbour-joining tree based on D_A distance; (b) neighbour-joining tree based on D_A distance with chimpanzees as the out-group population; (c) neighbour-joining tree based on D_{SW} distance with chimpanzees as the out-group population. Bootstrap values, indicating the degree of support for each branch point, are shown as the percentage of all replicates consistent with each branch point.

Evolutionary relationships

To assess the evolutionary relationships of the 16 populations, using the allele frequency data at the 23 loci we have constructed a neighbour-joining tree (Saitou and Nei 1987) based on the D_A distance measure (Nei *et al.* 1983). Although not proportional to evolutionary time, D_A is an efficient and robust measure suitable for the reconstruction of the correct phylogenetic tree using loci evolving via either infinite-allele or stepwise mutational mechanisms (Takezaki and Nei 1996; Rao *et al.* 1997). The neighbour-joining method is used for dendrogram construction because, unlike the UPGMA method, it does not assume that the rate of evolution in human populations is constant (Livshits and Nei 1990). The tree shown in figure 1a has several interesting features. First, populations within a major group cluster together. For example, the Asian populations (with the exception of the North Indian Brahmins) together with the American Indians form one cluster, so do the Caucasians and the African populations. Second, as evident from the bootstrap values, the statistical supports for these clusters are strong. Third, when rooted by the mid-point of the longest branch, the tree shows that the first separation of human populations occurred between the Africans and all the other, non-African populations.

Another way of rooting phylogenetic trees is by using an out-group population. Nei and Takezaki (1996) have shown that the root can be determined by using the chimpanzee as an out-group population. We have analysed an unrelated sample of chimpanzees at the same microsatellite loci [data not shown; allele frequencies at a subset of loci are presented elsewhere (Deka *et al.* 1994, 1995a)]. Interestingly, the chimpanzees are polymorphic at all but one locus, and at several loci the level of polymorphism is comparable to that of humans in terms of both size and number of alleles. The neighbour-joining tree based on the D_A distance measure, shown in figure 1b, with the chimpanzees as the out-group population, places the root between the African populations as a whole and all the other non-African populations. This placement of the root reaffirms the divergence of the African populations before the formation of the other populations. The bootstrap value of 99% shows very strong statistical support that all the non-African populations cluster together.

Although the D_A trees provide accurate topologies, they result in longer branch lengths for recently diverged

populations and shorter branch lengths for populations that diverged earlier. This is because D_A is not proportional to evolutionary time. On the other hand, the stepwise-weighted genetic distance D_{SW} is linear with respect to time, and therefore provides better estimates of branch lengths (Shriver *et al.* 1995). The D_{SW} tree shown in figure 1c also validates the hypothesis of the African ancestry of modern human populations. Branch lengths of the recently diverged populations are also relatively shorter in this tree. There are, however, some apparent inconsistencies in the topologies; for example, the location of the American Indians in the tree does not agree with their known phylogenetic position.

The positions of the two Asian Indian populations, viz. the Brahmin and the Kachari, in the phylogenetic trees are worth noting. The North Indian Brahmins belong to a cluster together with three other geographically diverse Caucasian populations (Brazilian White, German, unrelated CEPH parents), showing their affinity with the Caucasian gene pool. The Kachari, a distinct Mongoloid group from northeast India, on the other hand, form a cluster with the other Asian (Chinese and Japanese) and the Amerindian populations, reaffirming their position within the Mongoloid gene pool.

A point to be noted in the context of the above phylogenetic analyses is that the Papua New Guinea Highlanders and the Cabecars have been excluded from the dendrograms. Inclusion of these two populations substantially distorted the known and expected relationships of populations in the trees (not shown). This is due primarily to the allelic distributions in these two populations, which are markedly different from those observed in the remaining populations. We have noted that they have several high-frequency alleles, a possible signature of past population bottleneck or effect of drift operating on these populations. Relative isolation of the New Guinea Highlanders reflected in reduced variability has been observed earlier (Nei and Roychoudhury 1993; Deka *et al.* 1995a). Similarly, the Amerindian Cabecar population of Costa Rica has a very reduced population size (Barrantes *et al.* 1990).

Gene diversity analysis

The results of the gene diversity analysis are summarized in table 2, in which the total gene diversity (H_T) is decomposed into gene diversity between groups (G_g; between Caucasians, Africans, Asians, American Indians and Pacific Islanders) and that between populations within each group ($G_{p(g)}$). As the groups of populations show somewhat different within-group gene diversity (H_G), such analyses were also repeated for each group of populations separately to estimate the within population gene diversity (H_P) and the coefficient of gene diversity between populations within groups ($G_{p(g)}$). Further, to examine whether the pattern of gene diversity depends on the repeat motifs of loci, in addition to the combined analyses of the 23 loci computations were also done for the eight dinucleotide, two trinucleotide (without any disease implications), nine tetranucleotide and four disease-causing trinucleotide loci separately.

Overall, at a global level the total diversity (H_T) for the pooled loci is 74.9%, of which 90.6% is attributable to genetic variation within populations (resulting in $H_P = 67.9\%$). This is in agreement with observations that, of the total diversity, the maximum genetic variation is accounted for by interindividual differences within a population (Deka *et al.* 1995b; Barbujani *et al.* 1997). For the pooled data, the coefficient of gene diversity between groups of populations ($G_g = 5.3\%$) is larger than that among populations within groups ($G_{p(g)} = 4.1\%$). However, the latter is not uniform among all groups of populations. The smaller and isolated populations, such as the American Indians and the Pacific Islanders, show a somewhat smaller within-population (H_P) gene diversity and correspondingly a larger gene diversity between populations (9.5% and 9.1%, respectively, as opposed to a $G_{p(g)}$ of 1.5% to 1.8% for the Caucasians, Africans and Asians). This relationship between within-population heterozygosity and gene diversity follows the predictions of the mutation–drift model of microsatellite variation (Jin and Chakraborty 1995).

In general, the patterns of decomposition of gene diversity are similar for all repeat motifs, as seen from the analysis of dinucleotide, trinucleotide, tetranucleotide, and disease-causing trinucleotide loci separately. While the repeat motif types apparently do not show significantly different patterns (at 5% level of significance), the disease-causing trinucleotides show a somewhat larger level of within-population gene diversity (75%) compared with the other loci (62.9% to 67.8%). This is consistent with the hypothesis that even within the normal allele size ranges the disease-causing trinucleotides have a higher mutation rate than the other groups of loci (Chakraborty *et al.* 1997).

Since the loci grouped by their repeat motif types did not show any detectably different patterns of gene diversity, the results of interlocus variation of gene diversity analyses for individual loci are summarized in figures 2 and 3. In figure 2, we show the distributions of the coefficient of gene diversity between populations within groups ($G_{p(g)}$ in per cent) in five panels (a through e) for the five groups. Of course, these distributions are dependent on the choice of loci as well as populations. The three larger groups of populations (Caucasians, Africans and Asian Mongoloids) show a narrow range of variation of gene diversity between populations ($G_{p(g)}$ not exceeding 4% for any locus), whereas for the two isolated and smaller groups (American Indians and Pacific Islanders) the distributions are quite dispersed [$G_{p(g)}$ varied from 1.7% to 22.4% for the American Indians and from 0.5% to 31.4% for the Pacific Islanders]. In part, as discussed later, these differences are contributed by our sampling of populations. Nevertheless, given the average gene diversity between populations (1.5% for the Africans to 9.5% for the American Indians), the shapes of the distributions are consistent with that expected under a mutation–drift model (Nei and Chakravarti 1977). Incidentally, the

Table 2. Apportionment of genetic variation in 16 populations based on 23 microsatellite loci.

Group of populations and locus type	Heterozygosity (%) ± SE			Gene diversity (%) ± SE	
	Total (H_T)	Within groups (H_G)	Within population (H_P)	Among groups (G_g)	Among populations within group ($G_{p(g)}$)
Caucasians (four populations)					
Di-	–	69.06 ± 6.23	67.86 ± 6.08	–	1.72 ± 0.31
Tri-	–	63.95 ± 5.83	65.24 ± 5.90	–	1.06 ± 0.20
Tetra-	–	74.46 ± 1.47	73.43 ± 1.39	–	1.79 ± 0.24
Disease-associated	–	78.18 ± 1.47	77.43 ± 1.45	–	0.97 ± 0.08
Pooled	–	72.61 ± 2.34	71.48 ± 2.29	–	1.55 ± 0.16
Africans (three populations)					
Di-	–	79.40 ± 1.69	78.26 ± 1.51	–	1.43 ± 0.32
Tri-	–	78.58 ± 1.27	77.20 ± 1.61	–	1.76 ± 0.45
Tetra-	–	77.03 ± 2.32	75.85 ± 2.26	–	1.53 ± 0.16
Disease-associated	–	84.19 ± 0.72	82.95 ± 0.98	–	1.45 ± 0.49
Pooled	–	79.23 ± 1.18	78.04 ± 1.15	–	1.50 ± 0.15
Asians (three populations)					
Di-	–	69.65 ± 2.35	68.39 ± 2.25	–	1.81 ± 0.30
Tri-	–	70.57 ± 9.16	69.37 ± 9.23	–	1.69 ± 0.32
Tetra-	–	68.96 ± 2.29	67.84 ± 2.20	–	1.52 ± 0.20
Disease-associated	–	76.43 ± 5.95	74.68 ± 6.04	–	2.28 ± 0.36
Pooled	–	70.64 ± 1.69	69.35 ± 1.65	–	1.82 ± 0.14
American Indians (three populations)					
Di-	–	55.73 ± 8.19	50.35 ± 7.28	–	9.66 ± 1.79
Tri-	–	51.93 ± 22.91	49.12 ± 21.68	–	5.41 ± 0.02
Tetra-	–	69.06 ± 3.11	63.04 ± 2.89	–	8.71 ± 1.70
Disease-associated	–	72.66 ± 3.14	63.56 ± 5.73	–	12.53 ± 4.23
Pooled	–	63.56 ± 3.71	57.51 ± 3.41	–	9.52 ± 1.20
Pacific Islander (three populations)					
Di-	–	65.81 ± 6.11	62.68 ± 5.54	–	4.76 ± 1.23
Tri-	–	57.38 ± 15.25	52.69 ± 12.41	–	6.17 ± 2.79
Tetra-	–	65.49 ± 3.51	57.09 ± 4.22	–	12.82 ± 3.72
Disease-associated	–	83.61 ± 2.57	75.49 ± 6.02	–	9.72 ± 4.44
Pooled	–	68.05 ± 3.09	61.85 ± 3.10	–	9.10 ± 1.78
Total (16 populations)					
Di-	72.19 ± 4.24	67.32 ± 4.24	65.66 ± 4.13	5.81 ± 0.74	3.24 ± 0.55
Tri-	69.44 ± 10.80	64.50 ± 11.30	62.88 ± 9.90	6.47 ± 0.67	2.98 ± 0.50
Tetra-	74.61 ± 2.05	71.27 ± 2.04	67.82 ± 1.95	4.45 ± 0.98	4.64 ± 0.79
Disease-associated	83.67 ± 1.07	78.99 ± 1.30	74.99 ± 1.71	5.63 ± 0.92	4.75 ± 0.71
Pooled	74.90 ± 1.96	70.65 ± 2.00	67.89 ± 1.66	5.30 ± 0.46	4.06 ± 0.40

largest gene diversity in all population groups was observed at the dinucleotide repeat locus FLT1 located at the human fms-related tyrosine kinase gene (Polymeropoulos *et al.* 1991). This locus has the lowest within-population diversity, accompanied by the largest gene diversity across populations. Most populations have a single predominant allele at this locus. Frequency of the 15-repeat allele exceeds 80% in nine populations (see appendix). While a balancing selection can explain such observations, the effect of low mutation rate at this locus and/or more restricted allele size constraints cannot be distinguished from such data.

To further examine the differences of levels of gene diversity among groups of populations and those between populations within groups, we present in figure 3 the interlocus variation of these two coefficients of gene diversity (G_g in panel a and $G_{p(g)}$ in panel b). As shown in table 2, the average coefficient of gene diversity between groups of populations (5.3%) is larger than that between populations within groups (4.1%), but this difference is statistically not significant. However, a comparison of the interlocus distribution of G_g (panel a) with that of $G_{p(g)}$ (panel b) shows that the between-group variation is indeed significantly larger than variation among populations within the groups. The Wilcoxon's signed rank test for paired samples (Sokal and Rohlf 1981) shows that the distribution in panel a is significantly shifted towards the right compared to the one shown in panel b at the level $P = 0.0149$. In other words, the 23 loci together show that the genetic variation among populations within the five major groups is statistically smaller than the variation among the groups. It is interesting, however, that some loci have a tendency to produce smaller levels of gene diversity

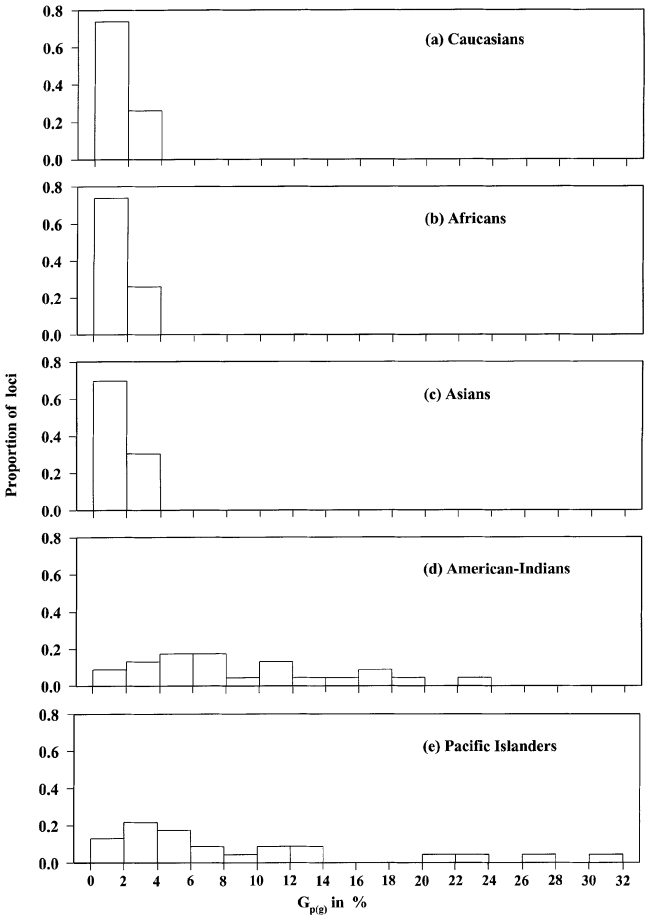

Figure 2. Interlocus distribution of coefficient of gene diversity among populations ($G_{p(g)}$ in per cent) within five major human population groups (shown separately in panels a–e) based on 23 microsatellite loci.

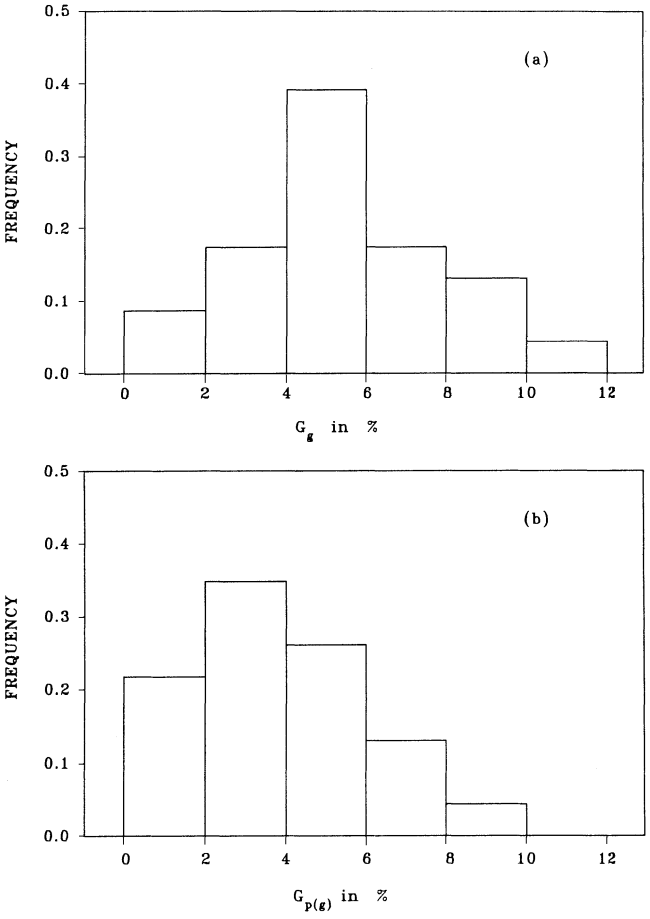

Figure 3. Interlocus distribution of coefficient of gene diversity in 16 populations: (a) gene diversity among the major groups of populations (G_g); (b) gene diversity among populations within the major population groups ($G_{p(g)}$).

across populations, as reflected in these computations. Whether or not this is due to homogenization of populations by forward–backward mutations of high rate, or due to more stringent allele size constraints at such loci, cannot be assessed from this analysis, and will be a subject for future study.

Discussion and conclusion

The goal of this study is to analyse the extent of genetic variation at a number of microsatellite loci in a sample of diverse human populations worldwide and to reassess the relationships of these populations in terms of the genetic affinity and diversity among them. With respect to the distribution of alleles at the 23 studied loci, although large variability exists among the examined populations, in general, populations within a continental group tend to show a higher degree of similarity. However, the American Indians and the Pacific Island populations show exceptions to this general observation. Historically it is known that the American Indians and the Pacific Islanders are isolated populations with smaller effective sizes. It is reasonable to speculate that population bottlenecks and genetic drift have been important factors in generating dramatic allele frequency differences in these populations and in extreme cases have resulted in near fixation of particular alleles in some populations, such as the New Guinea Highlander and the Cabecar. Furthermore, the three American Indian populations belong to different ethnic and linguistic stocks; the Dogrib are members of the Na-Dene groups, the Pehuenche Indians are Amerinds, and the Cabecars are linguistically Chibcha speakers. Similarly, the Pacific Island Samoans and New Guinea Highlanders are distinctly different populations except for being located in closer geographic proximity. As discussed below, the impact of such nonrandom sampling is also reflected in the extent of gene diversities among populations.

Phylogenetic analyses based on allele frequencies at 23 microsatellite loci show that the populations of African descent are the most diverse. We have employed two different genetic distance measures to reconstruct the phylogenetic

relationships of the examined populations. In spite of their inherent limitations, both the D_A and the D_{SW} trees indicate that the first separation of present-day humans took place between the Africans and all of the non-African populations. It adds to the already existent genetic evidence lending support to an African origin of all modern human populations (e.g. Cann *et al.* 1987; Bowcock *et al.* 1994; Nei 1995; Tishkoff *et al.* 1996; Shriver *et al.* 1997).

The gene diversity analyses show that the coefficient of gene diversity estimated from the 23 microsatellite loci is, in general, larger for populations that have remained isolated and probably small (such as the American Indians and Pacific Islanders; see table 2), among which the interlocus variation of gene diversity is also larger (figure 2). Further, the 23 loci in aggregate also show that the component of total gene diversity that can be ascribed to variation between groups of populations is significantly larger than that among populations within each major group. However, the absolute values of coefficient of gene diversity are to be interpreted with some caution, as these are dependent on the choice of loci as well as populations. Gene diversities between populations within the smaller groups (American Indians and Pacific Islanders) are probably overestimated because of their nonrandom sampling (noted above). Nevertheless, the inverse relationship between within-population gene diversity (H_P) and the coefficient of gene diversity between populations within groups ($G_{p(g)}$), namely that $G_{p(g)}$ is smaller as H_P becomes larger, is consistent with the mutation–drift expectations of gene diversity under the stepwise mutation model (Jin and Chakraborty 1995; Kimmel *et al.* 1996).

In the context of the objectives of this study, a brief discussion on the selection of loci should also be in place. In particular, all of the eight dinucleotide repeat loci are located on human chromosome 13, and therefore they are a nonrandom selection of dinucleotide repeats in the genome. We should like to note that originally these eight loci were selected from the same chromosome to examine how chromosomal linkage affects genotypic dependence between loci in unrelated individuals within populations (Deka *et al.* 1995a). This study demonstrated that, in spite of the presence of allelic association among a few closely linked loci (located within a genetic distance of 7 cM), these syntenic loci adequately revealed expected evolutionary relationships of the studied populations. Secondly, we should also like to comment on the inclusion of the four trinucleotide repeats which have been implicated in human diseases. We have noted that in terms of both number of alleles and heterozygosity these loci demonstrate a greater diversity. However, we have not observed any expanded alleles in the disease-causing range. Therefore the variation observed at these loci is representative of the total genomic variation reflective of the evolutionary histories of the examined populations. It is clearly evident from the distribution of allele frequencies at these loci (see appendix) that the populations of smaller effective sizes (such as the American Indians and the Pacific Islanders) have generally reduced levels of variability even at these loci compared to the other populations, although these loci are under some degree of selective pressure because of their genomic location and functional constraints. On the same token, at least seven other loci in this study, viz. FLT1, CSF1R, TH01, PLA2A1, F13A1, CYP19 and LPL, are located in or around functional genes. Further, although most microsatellite loci are expected to be located in the noncoding regions of the genomes, it is not unequivocally known which of them are associated with genes and which are not. Taking all of these facts into consideration, at this time it is reasonable to state that studies of human genetic variation and phylogenetic relationships using microsatellites or other DNA markers are not necessarily based on a complete set of neutral loci. It is therefore not surprising that the observed relationships of human populations based on such markers and traditional serum protein loci are in general congruent. Long-term evolutionary history is a product of mutation, selection and drift operating on these diverse sets of genomic sequences, and their study in conjunction will be essential in assessing the past relationships of contemporary humans accurately.

It is now well established that the mutational mechanisms at microsatellites are different from those of traditional serum protein markers. There is empirical evidence that mutations at microsatellite loci cause contraction as well as expansion of allele size (Weber and Wong 1993). A stepwise mutation–drift model has been advocated to account for this phenomenon (Shriver *et al.* 1993; Di Rienzo *et al.* 1994). Thus, genetic variation at these loci described by frequency distribution of alleles distinguished by their repeat lengths could be affected by the convergent features of the forward–backward stepwise mutational process, mathematical details of which are discussed elsewhere (e.g. see Kimmel and Chakraborty 1996; Kimmel *et al.* 1996). However, it has been shown that contraction or expansion bias of stepwise mutations does not affect the expectations of within-population heterozygosity or gene diversity between populations. Finally, the empirical data presented in this work and their analyses are in agreement with such theoretical predictions and reaffirm that examination of the major features of microevolutionary divergence of human populations is not compromised by these characteristics of the microsatellite loci.

Acknowledgements

We are thankful to Drs Emöke Szathmáry, Francisco Rothhammer, Peter Smouse, Jeff Long and Gebhard Flatz who generously provided DNA samples from the Dogrib Indian, Pehuenche Indian, New Guinea Highlander and German populations, respectively. We thank Diane Smelser for assisting in the preparation of the manuscript. This research was funded by grants GM45861, GM41399, GM58545 from the National Institutes of Health, SBR9600910 from the National Science Foundation, and 95-IJ-CX-008 and 97-LB-VX-0009 from the National Institute of Justice.

Appendix. Allele frequencies at 23 microsatellite loci in 16 human populations.

Locus		Populations*															
Fragment size (bp)	Repeat size	NI	BE	BB	GR	CP	BR	BW	CH	JP	KA	DG	PH	CA	AS	WS	NG
D13S71																	
65	12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
67	13	3.6	8.0	2.0	8.7	9.4	0.0	8.0	17.4	16.7	12.3	0.0	0.6	0.0	26.0	13.9	9.1
69	14	10.5	7.0	6.0	0.0	0.0	0.0	2.0	1.0	6.3	0.0	0.0	0.6	1.9	0.0	0.8	0.0
70	14.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0
71	15	5.9	5.0	2.0	0.0	0.0	2.8	0.0	2.0	0.0	0.9	3.7	0.6	1.9	0.0	0.0	0.0
72	15.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0
73	16	19.6	22.0	18.0	33.7	27.5	20.8	32.0	18.4	29.2	9.4	5.9	24.1	13.0	40.0	44.3	61.9
74	16.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.8	0.0	0.0	0.0	0.0
75	17	41.4	39.0	48.0	32.7	37.0	38.9	42.0	44.9	33.3	37.7	77.2	65.9	66.7	13.0	18.0	15.1
77	18	13.2	13.0	22.0	19.9	19.6	25.0	12.0	13.3	10.4	35.9	0.7	6.5	1.7	15.0	17.2	6.6
79	19	4.6	3.0	2.0	5.1	6.5	12.5	4.0	1.0	4.2	1.9	2.2	0.0	0.0	2.0	4.9	7.2
81	20	1.4	3.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.9	8.8	0.0	0.0	4.0	0.8	0.0
# Chr		220	100	50	196	138	72	50	98	48	106	136	170	54	100	122	318
D13S118																	
184	14	4.8	5.6	0.0	0.0	0.0	0.0	2.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
186	15	3.5	0.0	2.6	1.6	3.9	0.0	4.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.9
188	16	15.8	16.7	10.5	10.4	13.0	23.6	2.2	11.1	0.0	15.7	6.9	2.9	0.0	0.0	0.0	0.0
190	17	46.9	46.3	26.3	22.4	23.4	37.5	32.6	46.3	42.0	44.1	26.9	50.0	0.0	64.9	64.8	75.0
192	18	3.5	5.6	13.2	0.0	2.0	0.0	2.2	0.0	4.0	1.0	4.6	1.9	1.9	1.1	0.0	4.9
194	19	4.0	0.0	23.7	42.7	44.2	15.3	37.0	29.6	26.0	16.7	30.0	11.5	37.0	13.8	13.1	2.6
196	20	3.5	5.6	7.9	3.1	1.3	2.8	2.2	0.0	12.0	1.0	9.2	10.6	0.0	5.3	11.5	0.0
198	21	14.9	16.7	10.5	18.2	11.7	20.8	13.0	13.0	6.0	17.7	21.5	23.1	57.4	10.6	6.6	6.9
200	22	3.1	3.7	5.3	1.6	0.7	0.0	4.4	0.0	8.0	3.9	0.8	0.0	0.0	4.3	4.1	1.6
202	23	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	1.9	0.0	0.0	0.0
204	24	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.9	0.0	0.0	0.0
# Chr		228	54	38	192	154	72	46	54	50	102	130	208	54	94	122	304

* NI (Nigerian), BE (Benin), BB (Brazilian Black), GR (German), CP (CEPH parents), BR (Brahmin), BW (Brazilian White), CH (Chinese), JP (Japanese), KA (Kachari), DG (Dogrib), PE (Pehuenche), CA (Cabecar), AS (American Samoan), WS (Western Samoan), NG (New Guinea Highlander).

Locus		Populations*															
Fragment size (bp)	Repeat size	NI	BE	BB	GR	CP	BR	BW	CH	JP	KA	DG	PH	CA	AS	WS	NG
D13S121																	
158	15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0
160	16	10.8	11.5	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
162	17	3.6	1.9	2.0	7.3	8.3	14.1	6.0	8.0	12.0	7.1	0.0	0.0	0.0	0.0	0.0	0.0
164	18	9.5	5.8	14.0	3.7	3.2	10.9	0.0	4.0	2.0	1.0	5.3	0.5	0.0	0.0	0.0	0.3
166	19	32.0	38.5	38.0	49.5	43.0	43.8	50.0	46.0	40.0	42.9	69.7	78.9	74.1	68.0	56.6	66.8
168	20	6.8	3.9	16.0	8.9	10.3	4.7	10.0	6.0	14.0	11.2	2.3	1.8	9.3	4.0	18.0	5.6
170	21	5.9	9.6	0.0	5.2	7.1	9.4	10.0	22.0	24.0	25.5	6.8	9.2	0.0	18.0	19.7	6.6
172	22	14.9	19.2	20.0	7.3	5.8	7.8	6.0	8.0	8.0	4.1	0.8	4.1	9.3	5.0	0.8	2.6
174	23	8.6	5.8	4.0	11.5	14.1	6.3	6.0	4.0	0.0	8.2	15.2	4.6	7.4	2.0	1.6	3.0
176	24	3.6	0.0	2.0	6.8	8.3	3.1	8.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	3.3	10.5
178	25	0.9	1.9	0.0	0.0	0.0	0.0	4.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.6
180	26	3.6	1.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
# Chr		222	52	50	192	156	64	50	50	50	98	132	218	54	100	122	304
D13S122																	
83	8	0.4	0.0	0.0	1.1	3.3	1.5	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
85	9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
87	10	4.0	0.0	6.0	15.4	12.3	27.3	14.0	3.7	0.0	14.0	0.0	3.6	1.9	11.2	12.7	0.0
89	11	3.1	1.9	4.0	0.0	0.0	3.0	0.0	5.6	0.0	1.0	0.0	0.5	0.0	0.0	0.0	7.6
91	12	2.2	0.0	2.0	5.0	4.6	0.0	4.0	1.9	20.0	7.0	0.0	0.0	9.6	4.1	6.8	0.0
93	13	15.9	9.3	14.0	14.3	9.1	18.2	4.0	13.0	10.0	11.0	0.0	0.0	0.0	12.2	5.9	4.1
95	14	6.2	7.4	12.0	36.8	35.1	15.2	36.0	44.4	44.0	44.0	11.1	56.4	34.6	22.4	22.9	9.0
97	15	6.6	11.1	24.0	2.2	7.1	3.0	4.0	9.3	0.0	2.0	28.6	11.4	32.7	12.2	10.2	0.0
99	16	13.7	13.0	6.0	2.2	3.9	0.0	6.0	0.0	2.0	0.0	0.0	0.0	17.3	5.1	1.7	0.0
101	17	21.7	25.9	4.0	4.4	4.6	7.6	14.0	1.9	4.0	5.0	11.9	19.1	1.9	0.0	5.1	7.9
103	18	18.6	24.1	16.0	11.5	9.1	13.6	10.0	9.3	6.0	6.0	42.9	7.3	1.9	11.2	16.9	10.7
105	19	6.2	7.4	8.0	4.4	7.1	9.1	0.0	5.6	12.0	0.0	1.6	0.0	0.0	15.3	9.3	25.5
107	20	0.9	0.0	4.0	1.1	3.3	0.0	6.0	3.7	2.0	4.0	4.0	1.8	0.0	3.1	6.8	20.7
109	21	0.0	0.0	0.0	1.1	0.6	0.0	0.0	0.0	0.0	5.0	0.0	0.0	0.0	2.0	0.8	10.3
111	22	0.4	0.0	0.0	0.5	0.0	1.5	0.0	1.9	0.0	0.0	0.0	0.0	0.0	1.0	0.8	2.4
# Chr		226	66	50	182	154	66	50	54	50	100	126	220	52	98	118	290

* NI (Nigerian), BE (Benin), BB (Brazilian Black), GR (German), CP (CEPH parents), BR (Brahmin), BW (Brazilian White), CH (Chinese), JP (Japanese), KA (Kachari), DG (Dogrib), PE (Pehuenche), CA (Cabecar), AS (American Samoan), WS (Western Samoan), NG (New Guinea Highlander).

Locus		Populations*															
Fragment size (bp)	Repeat size	NI	BE	BB	GR	CP	BR	BW	CH	JP	KA	DG	PH	CA	AS	WS	NG
D13S124																	
173	9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
175	10	0.0	0.0	2.1	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
177	11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.2	6.6	0.0
179	12	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
181	13	4.3	1.0	0.0	0.0	0.6	0.0	2.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
183	14	18.0	21.0	29.2	0.5	0.0	2.8	4.0	25.5	13.6	10.4	0.0	0.0	0.0	36.5	34.4	6.9
185	15	47.1	42.0	31.3	46.9	44.2	36.1	44.0	16.3	15.9	29.3	8.8	15.5	14.8	45.8	45.9	38.4
187	16	18.4	25.0	25.0	41.8	35.3	55.6	42.0	53.1	59.1	56.6	88.2	83.9	33.3	12.5	13.1	54.7
189	17	3.2	8.0	0.0	1.5	3.2	1.4	0.0	1.0	2.3	0.0	2.9	0.0	29.6	0.0	0.0	0.0
191	18	6.1	2.0	10.4	5.1	10.3	4.2	6.0	0.0	0.0	0.9	0.0	0.6	22.2	0.0	0.0	0.0
193	19	2.2	0.0	0.0	4.1	6.4	0.0	2.0	0.0	9.1	2.8	0.0	0.0	0.0	0.0	0.0	0.0
195	20	0.0	1.0	2.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
# Chr		278	100	48	196	156	72	50	98	44	106	136	168	54	96	122	318
D13S193																	
127	12	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
129	13	8.0	11.5	12.0	1.1	4.6	0.0	0.0	0.0	0.0	1.3	0.0	0.9	1.9	0.0	0.0	0.0
131	14	23.6	23.1	24.0	17.0	15.1	28.6	18.0	7.4	8.0	21.8	26.5	23.8	22.2	34.0	29.5	14.3
133	15	35.4	34.6	32.0	13.3	13.2	17.1	16.0	51.9	46.0	42.3	32.6	17.3	7.4	21.0	29.5	54.5
135	16	12.7	13.5	12.0	0.0	0.7	1.4	0.0	0.0	2.0	2.6	6.1	0.0	0.0	0.0	0.0	5.2
137	17	2.8	1.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
139	18	0.9	0.0	0.0	1.1	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
141	19	0.5	0.0	0.0	2.1	2.6	0.0	2.0	0.0	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0
143	20	0.0	0.0	0.0	3.7	1.3	1.4	0.0	0.0	2.0	1.3	0.0	0.5	7.4	8.0	3.3	2.8
145	21	2.4	3.9	4.0	6.4	5.3	12.9	10.0	16.7	10.0	5.1	18.9	16.4	42.6	9.0	11.5	21.3
146	21.1	0.0	0.0	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
147	22	10.9	11.5	14.0	41.0	45.4	28.6	42.0	24.1	26.0	23.1	14.4	37.9	1.9	13.0	7.4	0.7
149	23	1.4	0.0	2.0	12.2	11.2	7.1	10.0	0.0	2.0	1.3	0.0	3.3	16.7	15.0	18.9	1.0
151	24	0.9	0.0	0.0	1.1	0.0	0.0	0.0	0.0	4.0	0.0	1.5	0.0	0.0	0.0	0.0	0.0
153	25	0.0	0.0	0.0	0.0	0.0	1.4	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
155	26	0.0	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
# Chr		212	52	50	188	152	70	50	54	50	78	132	214	54	100	122	286

* NI (Nigerian), BE (Benin), BB (Brazilian Black), GR (German), CP (CEPH parents), BR (Brahmin), BW (Brazilian White), CH (Chinese), JP (Japanese), KA (Kachari), DG (Dogrib), PE (Pehuenche), CA (Cabecar), AS (American Samoan), WS (Western Samoan), NG (New Guinea Highlander).

Locus: Fragment size (bp)	Repeat size	Populations*															
		NI	BE	BB	GR	CP	BR	BW	CH	JP	KA	DG	PH	CA	AS	WS	NG
D13S197																	
87		0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
97		0.0	0.0	0.0	11.7	16.2	8.6	10.0	0.0	0.0	1.0	2.3	0.0	0.0	0.0	0.9	6.5
98		0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
99		0.0	0.0	0.0	0.5	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.9	0.0	0.0	0.0
101		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.9	1.0	0.0	0.0	0.0	0.0	0.0	0.0
112		0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
118		3.2	0.0	10.9	0.5	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
119		0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
120		8.3	14.8	2.2	0.0	0.7	2.9	0.0	0.0	1.9	0.0	0.0	0.0	0.0	0.0	1.7	0.4
121		1.4	0.0	0.0	3.2	1.3	0.0	0.0	0.0	0.0	2.0	7.6	0.0	0.0	0.0	0.0	0.0
122		27.8	27.8	32.6	31.9	23.4	35.7	40.0	44.2	36.5	41.2	27.3	61.4	82.7	39.8	38.1	0.4
123		1.9	1.9	0.0	0.0	9.1	0.0	0.0	0.0	0.0	0.0	8.3	0.0	0.0	0.0	0.0	0.0
124		25.0	14.8	17.4	31.9	17.5	37.1	26.0	42.3	26.9	29.4	35.6	25.9	0.0	34.7	35.6	14.4
125		0.9	5.6	0.0	0.0	5.2	0.0	0.0	1.9	7.7	0.0	6.1	0.0	13.5	0.0	0.0	0.0
126		12.0	13.0	6.5	2.7	5.8	4.3	6.0	1.9	0.0	12.8	12.9	11.8	0.0	0.0	0.0	4.7
127		3.2	0.0	0.0	0.0	0.7	1.4	2.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
128		2.8	5.6	2.2	3.7	5.2	8.6	6.0	1.9	0.0	2.0	0.0	0.0	0.0	0.0	0.9	0.4
129		0.0	0.0	4.4	0.0	1.3	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
130		0.5	1.9	8.7	1.6	0.7	0.0	0.0	3.9	0.0	0.0	0.0	0.0	0.0	6.1	5.9	0.4
131		3.7	3.7	2.2	0.0	0.0	0.0	2.0	0.0	0.0	3.9	0.0	0.0	0.0	2.0	2.5	0.0
132		2.8	1.9	2.2	2.7	2.6	0.0	4.0	1.9	0.0	2.9	0.0	0.9	0.0	9.2	9.3	11.9
133		1.9	5.6	0.0	1.1	0.7	1.4	0.0	0.0	0.0	1.0	0.0	0.0	0.0	2.0	1.7	0.0
134		2.3	1.9	8.7	2.1	1.3	0.0	2.0	1.9	19.2	2.0	0.0	0.0	1.9	5.1	3.4	34.9
135		1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
136		0.5	1.9	0.0	0.0	0.7	0.0	0.0	0.0	3.9	0.0	0.0	0.0	0.0	0.0	0.0	26.3
138		0.0	0.0	0.0	0.0	3.9	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
139		0.0	0.0	0.0	4.8	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
140		0.0	0.0	2.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
142		0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
145		0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
# Chr		216	54	46	188	154	70	50	52	52	102	132	220	52	98	118	278

* NI (Nigerian), BE (Benin), BB (Brazilian Black), GR (German), CP (CEPH parents), BR (Brahmin), BW (Brazilian White), CH (Chinese), JP (Japanese), KA (Kachari), DG (Dogrib), PE (Pehuenche), CA (Cabecar), AS (American Samoan), WS (Western Samoan), NG (New Guinea Highlander).

Locus: Fragment size (bp)	Repeat size	Populations* NI	BE	BB	GR	CP	BR	BW	CH	JP	KA	DG	PH	CA	AS	WS	NG
FLT1																	
156	9	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
166	14	2.6	1.9	6.0	3.2	0.7	0.0	2.0	1.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7
167	14.1	0.0	0.0	0.0	0.5	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
168	15	38.9	35.2	46.0	83.7	89.0	80.6	84.0	53.7	67.3	55.9	89.2	94.4	100.0	81.0	80.3	71.0
170	16	4.3	1.9	4.0	3.7	3.3	9.7	0.0	1.9	1.9	5.9	0.0	0.0	0.0	2.0	0.0	2.8
172	17	6.0	9.3	2.0	1.6	0.0	0.0	0.0	0.0	3.9	0.0	5.4	0.0	0.0	0.0	0.0	0.0
174	18	9.0	11.1	12.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	3.2	0.0	0.0	0.0	0.0
176	19	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
178	20	3.4	3.7	6.0	0.0	0.7	0.0	0.0	5.6	1.9	4.9	0.0	0.0	0.0	0.0	0.0	0.0
179	20.1	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
180	21	4.7	5.6	2.0	0.0	0.0	0.0	2.0	1.9	0.0	1.0	5.4	0.0	0.0	0.0	0.8	1.7
182	22	9.0	20.4	6.0	5.8	4.6	9.7	8.0	31.5	21.2	26.5	0.0	2.3	0.0	17.0	18.9	23.1
184	23	13.3	7.4	6.0	0.5	0.0	0.0	0.0	3.7	3.9	2.9	0.0	0.0	0.0	0.0	0.0	0.7
186	24	7.3	3.7	8.0	0.5	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
188	25	0.4	0.0	0.0	0.5	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
190	26	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
200	27	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0
# Chr		234	54	50	190	154	72	50	54	52	102	130	216	54	100	122	290
PLA2A																	
115	9	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
118	10	0.0	0.0	0.0	1.0	2.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
121	11	4.8	1.0	15.7	46.1	44.9	54.3	37.9	28.4	27.1	23.0	17.0	28.4	4.9	28.0	31.0	80.4
124	12	2.9	9.2	12.9	12.8	13.3	12.9	16.1	2.0	0.0	8.0	2.0	2.0	5.9	11.0	9.0	12.7
127	13	18.3	13.3	8.6	2.0	2.0	0.0	3.2	19.6	27.1	9.0	15.0	15.7	10.8	19.0	17.0	0.0
130	14	28.9	23.5	17.1	14.7	14.3	11.4	23.4	28.4	26.0	25.0	30.0	21.6	7.8	16.0	12.0	4.9
133	15	26.0	28.6	28.6	15.7	19.4	14.3	13.7	12.8	7.3	18.0	31.0	28.4	60.8	14.0	16.0	2.0
136	16	19.2	24.5	17.1	6.9	4.1	5.7	4.8	6.9	9.4	17.0	5.0	3.9	8.8	12.0	15.0	0.0
139	17	0.0	0.0	0.0	1.0	0.0	1.4	0.0	2.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
# Chr		104	98	70	102	98	70	124	102	96	100	100	102	102	100	100	102

* NI (Nigerian), BE (Benin), BB (Brazilian Black), GR (German), CP (CEPH parents), BR (Brahmin), BW (Brazilian White), CH (Chinese), JP (Japanese), KA (Kachari), DG (Dogrib), PE (Pehuenche), CA (Cabecar), AS (American Samoan), WS (Western Samoan), NG (New Guinea Highlander).

Locus		Populations*															
Fragment size (bp)	Repeat size	NI	BE	BB	GR	CP	BR	BW	CH	JP	KA	DG	PH	CA	AS	WS	NG
D20S473																	
166	7	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
169	8	1.7	0.0	0.0	6.7	5.0	0.0	3.7	0.0	3.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
172	9	0.0	1.7	3.3	0.0	0.0	2.2	0.0	0.0	0.0	0.0	0.0	7.7	0.0	0.0	0.0	0.0
175	10	21.7	20.0	13.3	0.0	0.0	2.2	3.7	1.9	0.0	1.7	0.0	3.9	0.0	0.0	0.0	1.7
178	11	30.0	15.0	30.0	28.3	15.0	30.4	25.9	13.0	20.7	19.0	1.7	3.9	7.4	7.1	10.0	6.9
181	12	36.7	25.0	36.7	56.7	60.0	52.2	57.4	61.1	46.6	63.8	95.0	69.2	87.0	67.9	70.0	84.5
184	13	6.7	23.3	10.0	6.7	20.0	13.0	9.3	9.3	19.0	15.5	3.3	13.5	5.6	25.0	20.0	0.0
187	14	1.7	13.3	6.7	1.7	0.0	0.0	0.0	14.8	10.3	0.0	0.0	1.9	0.0	0.0	0.0	6.9
190	15	0.0	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
# Chr		60	60	60	60	60	46	54	54	58	60	60	52	54	28	30	58
TH01																	
180	5	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
184	6	12.8	14.3	12.5	20.6	18.4	24.3	19.5	15.7	26.0	15.0	2.0	34.3	67.7	5.0	9.0	16.7
188	7	52.9	57.1	48.4	15.7	23.5	17.1	21.2	19.6	27.1	30.0	54.0	33.3	13.7	55.0	40.0	0.0
192	8	15.7	20.4	21.9	8.8	9.2	25.7	10.2	2.9	4.2	7.0	6.0	0.0	0.0	12.0	21.0	78.4
196	9	13.7	5.1	4.7	16.7	8.2	20.0	14.4	52.0	40.6	40.0	31.0	2.0	1.0	7.0	3.0	3.9
199	9.3	4.9	2.0	12.5	36.3	39.8	12.9	34.8	3.9	1.0	8.0	7.0	30.4	17.7	21.0	26.0	1.0
200	10	0.0	0.0	0.0	2.0	0.0	0.0	0.0	5.9	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
204	11	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
# Chr		102	98	64	102	98	70	118	102	96	100	100	102	102	100	100	102

* NI (Nigerian), BE (Benin), BB (Brazilian Black), GR (German), CP (CEPH parents), BR (Brahmin), BW (Brazilian White), CH (Chinese), JP (Japanese), KA (Kachari), DG (Dogrib), PE (Pehuenche), CA (Cabecar), AS (American Samoan), WS (Western Samoan), NG (New Guinea Highlander).

Locus: Fragment size (bp)	Repeat size	Populations*															
		NI	BE	BB	GR	CP	BR	BW	CH	JP	KA	DG	PH	CA	AS	WS	NG
CSF1R																	
291	6	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
295	7	9.8	6.1	1.5	0.0	0.0	0.0	0.0	1.0	2.0	0.0	0.0	1.0	2.0	0.0	0.0	0.0
299	8	7.8	3.1	4.4	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0
303	9	4.9	4.1	1.5	2.0	0.0	7.6	3.4	3.9	3.1	5.1	9.0	3.9	0.0	0.0	1.0	2.5
307	10	30.4	31.6	25.0	25.5	32.7	21.2	22.0	17.7	22.9	22.5	39.0	27.5	13.7	16.0	13.0	15.0
311	11	16.7	23.5	42.7	33.3	25.5	24.2	32.2	26.5	16.7	36.7	14.0	19.6	30.4	42.0	39.0	38.8
315	12	28.4	25.5	19.1	28.4	34.7	36.4	38.1	40.2	46.9	29.6	29.0	41.2	25.5	33.0	35.0	42.5
319	13	1.0	5.1	4.4	9.8	5.1	10.6	1.7	7.8	7.3	6.1	6.0	5.9	26.5	8.0	10.0	1.3
323	14	0.0	0.0	1.5	0.0	1.0	0.0	1.7	2.9	0.0	0.0	0.0	1.0	0.0	0.0	2.0	0.0
327	15	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	1.0	0.0	3.0	0.0	0.0	1.0	0.0	0.0
# Chr		102	98	68	102	98	66	118	102	96	98	100	102	102	100	100	80
F13A1																	
281	3.2	26.9	7.0	19.1	6.1	4.3	11.1	9.0	34.3	40.4	23.3	48.0	33.0	48.0	7.0	6.0	3.6
283	4	7.7	7.0	7.1	6.1	2.9	5.6	4.0	13.7	6.4	15.6	8.0	18.0	18.6	3.0	2.0	0.0
287	5	30.8	40.0	25.0	18.4	11.4	34.7	20.0	10.8	7.5	8.9	17.0	3.0	30.4	26.0	20.0	90.9
291	6	14.1	5.0	14.3	26.5	34.3	22.2	28.0	41.2	45.7	34.4	24.0	11.0	1.0	61.0	69.0	5.5
295	7	12.8	16.0	19.1	34.7	40.0	20.8	32.0	0.0	0.0	10.0	2.0	35.0	2.0	2.0	3.0	0.0
299	8	5.1	8.0	8.3	1.0	1.4	0.0	0.0	0.0	0.0	1.1	1.0	0.0	0.0	0.0	0.0	0.0
303	9	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
307	10	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
311	11	1.3	2.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
315	12	1.3	3.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0
319	13	0.0	5.0	3.6	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
323	14	0.0	2.0	1.2	5.1	1.4	1.4	1.0	0.0	0.0	3.3	0.0	0.0	0.0	1.0	0.0	0.0
327	15	0.0	2.0	1.2	1.0	4.3	4.2	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
331	16	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	2.2	0.0	0.0	0.0	0.0	0.0	0.0
# Chr		78	100	84	98	70	72	100	102	94	90	100	100	102	100	100	110

* NI (Nigerian), BE (Benin), BB (Brazilian Black), GR (German), CP (CEPH parents), BR (Brahmin), BW (Brazilian White), CH (Chinese), JP (Japanese), KA (Kachari), DG (Dogrib), PE (Pehuenche), CA (Cabecar), AS (American Samoan), WS (Western Samoan), NG (New Guinea Highlander).

Locus: Fragment size (bp)	Repeat size	Populations*															
		NI	BE	BB	GR	CP	BR	BW	CH	JP	KA	DG	PH	CA	AS	WS	NG
CYP19																	
173	5	26.7	30.0	34.5	37.8	37.5	27.8	28.6	34.3	39.6	20.8	38.0	69.2	25.5	53.0	61.0	11.0
177	6	41.4	56.0	44.1	15.6	19.4	27.8	24.5	22.6	29.2	40.6	37.0	6.4	65.7	7.0	8.0	80.5
178	6.1	0.0	0.0	4.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
181	7	8.6	1.0	3.6	7.8	11.1	8.3	6.1	0.0	0.0	2.1	0.0	0.0	0.0	5.0	4.0	0.8
185	8	0.0	0.0	0.0	0.0	1.4	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
189	9	0.9	0.0	1.2	1.1	0.0	1.4	4.1	1.0	1.0	2.1	4.0	0.0	0.0	0.0	0.0	0.0
193	10	15.5	9.0	6.0	35.6	29.2	31.9	34.7	33.3	26.0	30.2	21.0	24.5	8.8	35.0	27.0	5.9
197	11	6.0	2.0	6.0	2.2	1.4	2.8	2.0	7.8	4.2	3.1	0.0	0.0	0.0	0.0	0.0	1.7
201	12	0.9	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
# Chr		116	100	84	90	72	72	98	102	78	96	100	94	102	100	100	118
LPL																	
115	8	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
119	9	10.0	23.9	10.7	2.4	4.2	1.4	4.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0
123	10	34.0	42.4	45.2	41.7	31.9	56.9	46.0	63.7	62.8	58.0	65.0	71.3	74.5	84.0	88.0	47.5
127	11	12.0	9.8	9.5	22.6	27.8	18.1	34.0	9.8	14.1	5.7	19.0	4.3	6.9	2.0	3.0	2.5
131	12	29.0	18.5	32.1	31.0	33.3	20.8	14.0	23.5	21.8	33.0	14.0	24.5	18.6	11.0	4.0	22.9
135	13	14.0	5.4	2.4	2.4	2.8	1.4	2.0	2.9	1.3	3.4	0.0	0.0	0.0	3.0	5.0	27.1
139	14	0.0	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
# Chr		100	92	84	84	72	72	100	102	78	88	100	94	102	100	100	118

* NI (Nigerian), BE (Benin), BB (Brazilian Black), GR (German), CP (CEPH parents), BR (Brahmin), BW (Brazilian White), CH (Chinese), JP (Japanese), KA (Kachari), DG (Dogrib), PE (Pehuenche), CA (Cabecar), AS (American Samoan), WS (Western Samoan), NG (New Guinea Highlander).

Locus Fragment size (bp)	Repeat size	Populations*															
		NI	BE	BB	GR	CP	BR	BW	CH	JP	KA	DG	PH	CA	AS	WS	NG
D20S604																	
115	8	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
119	9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.7	0.0	0.0	0.0	0.0	0.0
123	10	10.0	10.0	5.0	0.0	0.0	0.0	1.9	5.4	0.0	0.0	0.0	0.0	0.0	3.6	3.3	0.0
127	11	13.3	0.0	15.0	0.0	1.7	2.2	1.9	8.9	12.1	6.7	0.0	1.9	0.0	0.0	0.0	6.9
131	12	5.0	3.3	6.7	1.7	1.7	8.7	5.6	0.0	8.6	8.3	30.0	15.4	0.0	7.1	3.3	29.3
135	13	11.7	10.0	15.0	10.0	8.3	10.9	11.1	3.6	3.5	5.0	6.7	13.5	0.0	14.3	6.7	8.6
139	14	18.3	28.3	13.3	16.7	31.7	23.9	29.6	28.6	44.8	33.3	23.3	25.0	68.5	21.4	46.7	22.4
143	15	25.0	23.3	21.7	56.7	40.0	37.0	31.5	37.5	15.5	31.7	10.0	28.9	29.6	35.7	33.3	29.3
147	16	13.3	23.3	18.3	13.3	13.3	13.0	18.5	12.5	12.1	13.3	28.3	15.4	1.9	14.3	6.7	3.4
151	17	1.7	1.7	5.0	1.7	3.3	4.4	0.0	3.6	3.5	1.7	0.0	0.0	0.0	3.6	0.0	0.0
# Chr		60	60	60	60	60	46	54	56	58	60	60	52	54	28	30	58
D20S481																	
213	9	0.0	0.0	0.0	0.0	0.0	0.0	1.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
217	10	1.7	0.0	0.0	3.3	0.0	2.2	1.9	0.0	0.0	1.7	0.0	0.0	0.0	0.0	0.0	0.0
221	11	6.7	1.7	3.3	3.3	5.0	0.0	3.7	0.0	0.0	3.3	0.0	1.9	1.9	3.6	0.0	8.6
225	12	11.7	18.3	15.0	8.3	11.7	2.2	3.7	5.4	0.0	3.3	0.0	0.0	0.0	3.6	0.0	22.4
229	13	3.3	0.0	6.7	5.0	5.0	0.0	5.6	1.8	1.7	1.7	1.7	0.0	0.0	0.0	0.0	12.1
233	14	3.3	5.0	16.7	21.7	25.0	10.9	18.5	1.8	1.7	5.0	35.0	34.6	66.7	17.9	6.7	17.2
237	15	26.7	16.7	15.0	18.3	11.7	23.9	20.3	12.5	5.2	21.7	21.7	7.7	7.4	21.4	0.0	13.8
241	16	23.3	23.3	23.3	10.0	26.7	39.1	24.1	42.9	39.7	45.0	26.7	30.8	1.9	42.9	80.0	15.5
245	17	16.7	23.3	18.3	18.3	8.3	19.6	14.8	33.9	43.1	16.7	10.0	23.1	22.2	10.7	13.3	8.6
249	18	6.7	11.7	1.7	10.0	6.7	2.2	5.6	1.8	6.9	1.7	5.0	1.9	0.0	0.0	0.0	1.7
253	19	0.0	0.0	0.0	1.7	0.0	0.0	0.0	0.0	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
# Chr		60	60	60	60	60	46	54	56	58	60	60	52	54	28	30	58

* NI (Nigerian), BE (Benin), BB (Brazilian Black), GR (German), CP (CEPH parents), BR (Brahmin), BW (Brazilian White), CH (Chinese), JP (Japanese), KA (Kachari), DG (Dogrib), PE (Pehuenche), CA (Cabecar), AS (American Samoan), WS (Western Samoan), NG (New Guinea Highlander).

Locus: Fragment size (bp)	Repeat size	Populations*															
		NI	BE	BB	GR	CP	BR	BW	CH	JP	KA	DG	PH	CA	AS	WS	NG
D21S1435																	
159	7	0.0	3.3	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
163	8	3.5	1.7	3.3	3.3	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
167	9	5.2	11.7	6.7	0.0	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
171	10	12.1	11.7	21.7	20.0	21.7	10.9	16.0	6.7	10.3	5.0	26.7	16.7	8.0	23.1	10.0	1.7
175	11	29.3	28.3	13.3	35.0	35.0	34.8	34.0	38.3	31.0	43.3	1.7	42.6	42.0	34.6	53.3	33.3
179	12	19.0	15.0	35.0	13.3	20.0	17.4	32.0	18.3	29.3	20.0	36.7	20.4	42.0	11.5	6.7	35.0
183	13	24.1	25.0	15.0	25.0	18.3	34.8	14.0	25.0	22.4	23.3	35.0	14.8	8.0	23.1	23.3	28.3
187	14	3.5	3.3	3.3	3.3	3.3	2.2	2.0	11.7	5.2	8.3	0.0	1.9	0.0	7.7	6.7	1.7
191	15	3.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.7	0.0	0.0	3.7	0.0	0.0	0.0	0.0
# Chr		58	60	60	60	60	46	50	60	50	60	60	54	50	26	30	60
D21S1446																	
201		8.6	8.3	13.0	0.0	0.0	2.1	0.0	0.0	0.0	0.0	0.0	1.9	0.0	0.0	0.0	0.0
205		0.0	1.7	1.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.7
209		25.9	18.3	14.8	45.0	36.7	43.8	46.0	55.2	51.7	60.0	34.5	33.3	39.6	53.9	70.0	48.3
211		0.0	3.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
213		6.9	5.0	13.0	15.0	10.0	10.4	16.0	20.7	6.9	13.3	1.7	7.4	0.0	34.6	26.7	23.3
215		1.7	0.0	0.0	3.3	1.7	2.1	0.0	3.5	0.0	0.0	0.0	0.0	4.2	0.0	0.0	0.0
217		0.0	3.3	9.3	11.7	8.3	4.2	10.0	0.0	1.7	1.7	6.9	9.3	0.0	0.0	0.0	1.7
219		15.5	18.3	9.3	0.0	3.3	4.2	0.0	0.0	1.7	3.3	27.6	1.9	0.0	0.0	0.0	0.0
221		5.2	6.7	3.7	3.3	6.7	0.0	4.0	0.0	0.0	0.0	0.0	0.0	16.7	0.0	0.0	0.0
223		36.2	31.7	33.3	18.3	31.7	27.1	24.0	20.7	31.0	20.0	29.3	46.3	39.6	11.5	3.3	25.0
225		0.0	0.0	0.0	0.0	0.0	2.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
227		0.0	3.3	1.9	3.3	1.7	4.2	0.0	0.0	6.9	1.7	0.0	0.0	0.0	0.0	0.0	0.0
229		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.7	0.0	0.0	0.0	0.0	0.0	0.0
# Chr		58	60	54	60	60	48	50	58	58	60	58	54	48	26	30	60

* NI (Nigerian), BE (Benin), BB (Brazilian Black), GR (German), CP (CEPH parents), BR (Brahmin), BW (Brazilian White), CH (Chinese), JP (Japanese), KA (Kachari), DG (Dogrib), PE (Pehuenche), CA (Cabecar), AS (American Samoan), WS (Western Samoan), NG (New Guinea Highlander).

Locus		Populations*															
Fragment size (bp)	Repeat size	NI	BE	BB	GR	CP	BR	BW	CH	JP	KA	DG	PH	CA	AS	WS	NG
DM																	
75	4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
78	5	26.4	32.0	19.1	32.7	40.0	33.3	39.0	29.4	19.0	30.2	31.0	6.5	0.0	31.6	35.0	0.0
81	6	0.0	2.0	0.0	0.0	0.0	2.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
84	7	2.8	2.0	2.4	0.0	0.0	0.0	1.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
87	8	1.9	3.0	3.6	0.0	1.3	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
90	9	3.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.1	4.0	8.0
93	10	11.3	7.0	2.4	0.0	2.5	1.4	3.0	1.0	1.0	12.8	2.0	8.7	0.0	0.0	0.0	4.0
96	11	17.9	15.0	11.9	9.6	13.1	19.4	10.0	6.9	18.0	16.3	26.0	45.7	5.0	4.1	9.0	34.0
99	12	17.9	16.0	17.9	12.5	11.9	8.3	10.0	26.5	22.0	18.6	15.0	3.3	12.0	2.0	9.0	12.0
102	13	8.5	9.0	22.6	15.4	14.4	13.9	24.0	14.7	20.0	11.6	21.0	26.1	83.0	29.6	28.0	23.0
105	14	5.7	12.0	13.1	6.7	3.8	4.2	2.0	8.8	2.0	5.8	0.0	1.1	0.0	9.2	6.0	12.0
108	15	2.8	0.0	0.0	0.0	0.0	4.2	3.0	4.9	6.0	0.0	0.0	0.0	0.0	18.4	8.0	1.0
111	16	0.0	1.0	1.2	1.0	5.6	5.6	0.0	2.0	4.0	1.2	0.0	2.2	0.0	0.0	0.0	1.0
114	17	0.0	0.0	0.0	1.0	0.0	2.8	0.0	0.0	0.0	1.2	0.0	0.0	0.0	0.0	0.0	3.0
117	18	0.0	1.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
120	19	0.9	0.0	0.0	2.9	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0
123	20	0.0	0.0	0.0	1.9	1.3	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
126	21	0.0	0.0	2.4	5.9	2.5	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
129	22	0.0	0.0	0.0	5.8	1.9	0.0	2.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
132	23	0.0	0.0	1.2	1.0	0.6	2.8	1.0	1.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
135	24	0.0	0.0	1.2	1.0	0.6	0.0	0.0	0.0	1.0	1.2	0.0	5.4	0.0	0.0	0.0	0.0
138	25	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
141	26	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0
144	27	0.0	0.0	0.0	1.9	0.0	1.4	0.0	0.0	0.0	0.0	2.0	1.1	0.0	0.0	0.0	0.0
147	28	0.0	0.0	1.2	0.0	0.6	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
159	32	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
162	33	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
165	34	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
# Chr		100	100	84	100	100	72	100	102	100	86	100	92	100	98	100	100

* NI (Nigerian), BE (Benin), BB (Brazilian Black), GR (German), CP (CEPH parents), BR (Brahmin), BW (Brazilian White), CH (Chinese), JP (Japanese), KA (Kachari), DG (Dogrib), PE (Pehuenche), CA (Cabecar), AS (American Samoan), WS (Western Samoan), NG (New Guinea Highlander).

Locus		Populations*															
Fragment size (bp)	Repeat size	NI	BE	BB	GR	CP	BR	BW	CH	JP	KA	DG	PH	CA	AS	WS	NG
SCA																	
184	19	0.0	0.0	3.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
190	21	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
193	22	0.0	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
196	23	0.0	0.0	2.5	0.0	0.0	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0
199	24	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	1.4	1.4	0.0	0.0	0.0	0.0	0.0	0.0
202	25	1.2	1.0	0.0	0.0	0.0	3.9	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
205	26	9.3	24.5	7.5	0.0	1.1	0.0	0.0	10.0	13.9	20.0	1.1	8.6	29.0	9.2	13.3	1.3
208	27	12.8	9.2	7.5	0.0	2.1	1.9	2.1	2.0	0.0	1.4	0.0	0.0	0.0	1.0	0.0	1.3
211	28	7.0	7.1	8.8	1.1	3.2	11.5	5.2	48.0	33.3	20.0	34.0	30.0	17.0	38.8	25.5	0.0
214	29	10.5	5.1	7.5	41.3	31.9	32.7	29.2	16.0	16.7	18.6	10.6	1.4	0.0	15.3	16.3	17.9
217	30	29.1	25.5	26.3	31.5	38.3	34.6	37.5	16.0	33.3	20.0	22.3	25.7	15.0	24.5	23.5	9.0
220	31	9.3	13.3	17.5	10.9	5.3	11.5	10.4	2.0	1.4	5.7	22.3	21.4	38.0	7.1	7.1	0.0
223	32	10.5	9.2	5.0	10.9	12.8	1.9	9.4	0.0	0.0	4.3	0.0	2.9	1.0	1.0	4.1	0.0
226	33	3.5	2.0	8.8	1.1	2.1	1.9	2.1	0.0	0.0	2.9	9.6	8.6	0.0	0.0	3.0	15.4
229	34	1.2	2.0	2.5	0.0	0.0	0.0	0.0	2.0	0.0	2.9	0.0	0.0	0.0	2.0	7.1	25.6
232	35	1.2	0.0	1.3	2.2	3.2	0.0	4.2	0.0	0.0	1.4	0.0	1.4	0.0	0.0	0.0	19.2
235	36	3.5	0.0	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0
238	37	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.3
250	41	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
# Chr		86	98	80	92	94	52	96	50	72	70	94	70	100	98	98	78

* NI (Nigerian), BE (Benin), BB (Brazilian Black), GR (German), CP (CEPH parents), BR (Brahmin), BW (Brazilian White), CH (Chinese), JP (Japanese), KA (Kachari), DG (Dogrib), PE (Pehuenche), CA (Cabecar), AS (American Samoan), WS (Western Samoan), NG (New Guinea Highlander).

Locus: Fragment size (bp)	Repeat size	Populations*															
		NI	BE	BB	GR	CP	BR	BW	CH	JP	KA	DG	PH	CA	AS	WS	NG
DRPLA																	
116	6	1.0	0.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	5.0
119	7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	2.0
122	8	4.1	3.0	7.1	10.0	16.3	11.8	12.0	0.0	1.1	7.1	0.0	1.1	0.0	18.4	14.0	16.0
125	9	14.3	13.0	3.6	0.0	0.0	0.0	0.0	2.0	1.1	0.0	0.0	0.0	0.0	2.0	6.0	5.0
128	10	1.0	1.0	1.2	9.0	7.6	7.4	5.0	17.0	18.9	10.7	67.4	0.0	6.3	15.3	9.0	42.0
131	11	6.1	5.0	4.8	0.0	0.0	0.0	1.0	0.0	6.7	0.0	0.0	7.8	0.0	6.1	5.0	2.0
134	12	24.5	26.0	22.6	1.0	1.1	7.4	4.0	1.0	2.2	9.5	0.0	3.3	1.0	2.0	4.0	0.0
137	13	32.7	32.0	13.1	1.0	1.1	0.0	9.0	3.0	3.3	3.6	5.1	7.8	12.5	6.1	7.0	3.0
140	14	6.1	6.0	7.1	9.0	3.3	13.2	3.0	1.0	3.3	3.6	0.0	1.1	14.6	11.2	7.0	4.0
143	15	6.1	12.0	31.0	38.0	37.0	35.3	40.0	28.0	21.1	31.0	24.5	74.4	61.5	9.2	12.0	8.0
146	16	1.0	1.0	7.1	24.0	18.5	20.6	17.0	1.0	4.4	16.7	3.1	3.3	4.2	7.1	9.0	0.0
149	17	0.0	1.0	1.2	5.0	6.5	2.9	4.0	23.0	15.6	6.0	0.0	1.1	0.0	9.2	9.0	0.0
152	18	3.1	0.0	0.0	0.0	2.2	1.5	2.0	15.0	13.3	8.3	0.0	0.0	0.0	1.0	4.0	12.0
155	19	0.0	0.0	0.0	2.0	1.1	0.0	2.0	5.0	6.7	3.6	0.0	0.0	0.0	1.0	1.0	1.0
158	20	0.0	0.0	0.0	1.0	3.3	0.0	0.0	1.0	1.1	0.0	0.0	0.0	0.0	3.1	6.0	0.0
161	21	0.0	0.0	0.0	0.0	2.2	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	5.1	4.0	0.0
164	22	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	2.0	0.0
167	23	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	0.0	0.0	0.0	0.0	0.0	1.0	0.0
170	24	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
176	26	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
# Chr		98	100	84	100	92	68	100	100	90	84	98	90	96	98	100	100

* NI (Nigerian), BE (Benin), BB (Brazilian Black), GR (German), CP (CEPH parents), BR (Brahmin), BW (Brazilian White), CH (Chinese), JP (Japanese), KA (Kachari), DG (Dogrib), PE (Pehuenche), CA (Cabecar), AS (American Samoan), WS (Western Samoan), NG (New Guinea Highlander).

Locus		Populations*															
Fragment size (bp)	Repeat size	NI	BE	BB	GR	CP	BR	BW	CH	JP	KA	DG	PH	CA	AS	WS	NG
HD																	
89	7	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
95	9	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
98	10	0.0	0.0	0.0	0.0	0.6	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
101	11	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
104	12	2.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.2	2.0	0.0
107	13	0.0	0.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
110	14	4.2	1.0	0.0	1.0	0.6	1.4	1.0	0.0	0.9	1.1	0.0	0.0	8.2	7.3	8.0	79.5
113	15	34.2	33.0	28.6	7.3	9.5	9.7	12.0	2.0	3.6	3.3	27.1	26.1	5.1	4.2	2.0	0.0
116	16	7.5	11.0	13.1	9.4	4.4	5.6	3.0	4.0	7.1	2.2	0.0	2.3	0.0	6.3	4.0	0.0
119	17	11.7	10.0	14.3	40.6	38.0	47.2	37.0	72.0	44.6	65.6	52.1	56.8	18.4	55.2	49.0	0.0
122	18	7.5	14.0	10.7	7.3	12.0	19.4	8.0	14.0	26.8	14.4	3.1	4.6	3.1	8.3	16.0	0.0
125	19	13.3	13.0	14.3	6.3	7.0	6.9	13.0	2.0	7.1	2.2	0.0	2.3	0.0	1.0	2.0	0.0
128	20	9.2	7.0	3.6	6.3	10.8	0.0	9.0	0.0	3.6	1.1	0.0	1.1	1.0	1.0	0.0	2.3
131	21	0.8	4.0	2.4	1.0	6.3	2.8	1.0	1.0	0.9	1.1	0.0	0.0	13.3	2.1	6.0	6.1
134	22	0.0	3.0	1.2	3.1	2.5	0.0	0.0	0.0	0.0	1.1	1.0	0.0	0.0	1.0	6.0	3.0
137	23	3.3	0.0	2.4	4.2	1.9	0.0	6.0	3.0	0.9	3.3	8.3	1.1	18.4	0.0	0.0	0.8
140	24	1.7	0.0	1.2	4.2	1.3	6.9	4.0	1.0	2.7	1.1	2.1	4.6	21.4	6.3	1.0	0.0
143	25	0.8	0.0	1.2	1.0	0.6	0.0	0.0	0.0	0.0	1.1	0.0	0.0	2.0	1.0	4.0	1.5
146	26	0.0	2.0	1.2	3.1	0.6	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	1.0	0.0	3.8
149	27	0.8	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	1.1	2.1	0.0	9.2	0.0	0.0	0.0
152	28	0.8	0.0	0.0	0.0	0.6	0.0	1.0	0.0	0.0	0.0	4.2	0.0	0.0	0.0	0.0	0.8
155	29	0.8	1.0	3.6	0.0	0.6	0.0	2.0	0.0	0.0	0.0	0.0	1.1	0.0	0.0	0.0	0.8
158	30	0.0	0.0	1.2	2.1	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.5
167	33	0.8	0.0	0.0	0.0	0.6	0.0	1.0	0.0	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0
170	34	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
# Chr		120	100	84	96	158	72	100	100	112	90	96	88	98	96	100	132

* NI (Nigerian), BE (Benin), BB (Brazilian Black), GR (German), CP (CEPH parents), BR (Brahmin), BW (Brazilian White), CH (Chinese), JP (Japanese), KA (Kachari), DG (Dogrib), PE (Pehuenche), CA (Cabecar), AS (American Samoan), WS (Western Samoan), NG (New Guinea Highlander).

References

Barbujani G., Magagni A., Minch E. and Cavalli-Sforza L. L. 1997 An apportionment of human DNA diversity. *Proc. Natl. Acad. Sci. USA* **94**, 4516–4519.

Barrantes R., Smouse P. E., Mohrenweiser H. W., Gershowitz H., Azofeifa J., Arias T. D. and Neel J. V. 1990 Microevolution in lower central America: genetic characterization of the Chibcha-speaking groups of Costa Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity. *Am. J. Hum. Genet.* **46**, 63–84.

Bowcock A. M., Ruiz-Linares A., Tomfohrde J., Minch E., Kidd J. R. and Cavalli-Sforza L. L. 1994 High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455–457.

Calafell F., Shuster A., Speed W. C., Kidd J. R. and Kidd K. K. 1998 Short tandem repeat polymorphism evolution in humans. *Eur. J. Hum. Genet.* **6**, 38–49.

Cann R. L., Sotneking M. and Wilson A. C. 1987 Mitochondrial DNA and human evolution. *Nature* **325**, 548–553.

Chakraborty R. 1980 Gene-diversity analysis in nested subdivided populations. *Genetics* **96**, 721–723.

Chakraborty R. and Jin L. 1993 Determination of relatedness between individuals by DNA fingerprinting. *Hum. Biol.* **65**, 875–895.

Chakraborty R., Kimmel M., Stivers D. N., Davison L. J. and Deka R. 1997 Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc. Natl. Acad. Sci. USA* **94**, 1041–1046.

Deka R., Shriver M. D., Yu L. M., Jin L., Aston C. E., Chakraborty R. and Ferrell R. E. 1994 Conservation of human chromosome 13 polymorphic microsatellite (CA)n repeats in chimpanzees. *Genomics* **22**, 226–230.

Deka R., Jin L., Shriver M. D., Yu L. M., DeCroo S., Hundrieser J., Bunker C. H., Ferrell R. E. and Chakraborty R. 1995a Population genetics of dinucleotide (dC-dA)n.(dG-dT)n polymorphisms in world populations. *Am. J. Hum. Genet.* **56**, 461–474.

Deka R., Shriver M. D., Yu L. M., Ferrell R. E. and Chakraborty R. 1995b Intra- and inter-population diversity at short tandem repeat loci in diverse populations of the world. *Electrophoresis* **16**, 1659–1664.

Deka R., Shriver M. D., Jin L., Yu L. M., Ferrell R. E. and Chakraborty R. 1998 Tracing the origin of modern humans using nuclear microsatellite polymorphisms. In *The origins and past of modern humans—towards reconciliation* (ed. K. Omoto and P. V. Tobias), pp. 3–15. World Scientific, Singapore.

Di Rienzo A., Peterson A. C., Garza J. C., Valdes A. M., Slatkin M. and Freimer N. B. 1994 Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl. Acad. Sci. USA* **91**, 3166–3170.

Jin L. and Chakraborty R. 1995 Population structure, stepwise mutations, heterozygote deficiency and their implications in DNA forensics. *Heredity* **74**, 274–285.

Jorde L. B., Bamshad M. J., Watkins W. S., Zenger R., Fraley A. E., Krakowiak P. A., Carpenter K. D., Soodyal H., Jenkins T. and Rogers A. R. 1995 Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. *Am. J. Hum. Genet.* **57**, 523–538.

Kimmel M. and Chakraborty R. 1996 Measures of variation at DNA repeat loci under a general stepwise mutation model. *Theor. Pop. Biol.* **50**, 345–367.

Kimmel M., Chakraborty R., Stivers D. N. and Deka R. 1996 Dynamics of repeat polymorphisms under a forward-backward mutation model: within- and between-population variability at microsatellite loci. *Genetics* **143**, 549–555.

Kimmel M., Chakraborty R., King J. P., Bamshad M., Watkins W. S. and Jorde L. B. 1998 Signature of population expansion in microsatellite repeat data. *Genetics* **148**, 1921–1930.

Li C. C. 1976 *First course in population genetics*. Boxwood, Pacific Grove, USA.

Livshits G. and Nei M. 1990 Relationship between intrapopulational and interpopulational genetic diversity in man. *Ann. Hum. Biol.* **17**, 501–513.

Nei M. 1973 Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* **70**, 3321–3323.

Nei M. 1995 Genetic support for the out-of-Africa theory of human evolution. *Proc. Natl. Acad. Sci. USA* **92**, 6720–6722.

Nei M. and Chakravarti A. 1977 Drift variances of F_{ST} and G_{ST} statistics obtained from a finite number of isolated populations. *Theor. Pop. Biol.* **11**, 307–325.

Nei M. and Roychoudhury A. K. 1993 Evolutionary relationships of human populations on a global scale. *Mol. Biol. Evol.* **10**, 927–943.

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

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

Pérez-Lezaun A., Calafell F., Mateu E., Comas D., Ruiz-Pacheco R. and Bertranpetit J. 1997 Microsatellite variation and the differentiation of modern humans. *Hum. Genet.* **99**, 1–7.

Polymeropoulos M. H., Rath D. S., Xiao H. and Merril C. R. 1991 Dinucleotide repeat polymorphism at the human fms-related tyrosine kinase gene (FLT1). *Nucl. Acids Res.* **19**, 2803.

Rao B. K., Sil S. B. and Majumder P. P. 1997 How useful are microsatellite loci in recovering short-term evolutionary history? *J. Genet.* **76**, 181–188.

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

Shriver M. D., Jin L., Chakraborty R. and Boerwinkle E. 1993 VNTR allele frequency distributions under the stepwise mutation model: A computer simulation approach. *Genetics* **134**, 983–993.

Shriver M. D., Jin L., Boerwinkle E., Deka R., Ferrell R. E. and Chakraborty R. 1995 A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol. Biol. Evol.* **12**, 914–920.

Shriver M. D., Jin L., Ferrell R. E. and Deka R. 1997 Microsatellite data support an early population expansion in Africa. *Genome Res.* **7**, 586–591.

Sokal R. R. and Rohlf F. J. 1981 *Biometry*, 2nd edition, p. 448. Freeman, New York.

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

Tishkoff S. A., Dietzsch E., Speed W., Pakstis A. J., Kidd J. R., Cheung K., Bonné-Tamir B., Santachiara-Benerecetti A. S., Moral P., Krings M., Pääbo S., Watson E., Risch N., Jenkins T. and Kidd K. K. 1996 Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* **271**, 1380–1387.

Weber J. L. and Wong C. 1993 Mutation of human short tandem repeats. *Hum. Mol. Genet.* **2**, 1123–1128.

Received 17 March 1999; in revised form 17 July 1999