Supplementary information for "Genetic structure of human populations"

Methods

Sample, markers, and genotypes: The data set that we analyzed differs from the HGDP-CEPH Human Genome Diversity Cell Line Panel of 1064 individuals in its inclusion of Japanese individual #1026, whose cell line could not be produced owing to technical problems, and its exclusions of She #1331, who was not genotyped, and 8 individuals whose populations had samples of size 1 or 2 (#993, #994, #1028, #1030, #1031, #1033, #1034, #1035). Individual #1410, who is not included in the Cell Line Panel, was genotyped, but as the only representative of his population, was not analyzed. The loci studied, from Marshfield Screening Set #10 (http://research.marshfieldclinic.org/ genetics/sets/combo.html), include a mixture of 377 polymorphic di-, tri-, and tetra-nucleotide repeat loci spread across all 22 autosomes (*2, 19*), with 3.8% missing data. Genotyping was performed by the Mammalian Genotyping Service (*19*).

"Africa" in this article refers to Sub-Saharan Africa, and "Middle East" includes the Mozabite population of Algeria. "Melanesian" is used in place of the usual name "Nasioi" (*29*), and "Colombian" includes individuals of multiple groups from Colombia. "Han (N. China)" includes Han individuals #1287-1296 sampled from northern China by the Chinese Human Genome Diversity Project. "Han" includes individuals born in China and sampled in the United States (San Francisco Bay area) by the laboratory of L. L. Cavalli-Sforza.

Analysis of variance: Variance components were estimated with GDA (*31*), assuming Hardy-Weinberg equilibrium within populations, and taking into account identity or non-identity of alleles but not allele sizes (*32*). Confidence intervals are based on 1000 bootstraps across loci. The World-B97 sample includes Bantu, Biaka, Cambodian, French, Han (N. China), Japanese, Karitiana, Mandenka, Maya, Mbuti, Melanesian, Papuan, Surui, and Tuscan.

Cluster analysis: All *structure* runs used 10,000 iterations after a burn-in of length 20,000, with a model of correlated allele frequencies (*14*). In highly structured data, as the number of clusters is increased, the most divergent groups typically separate into distinct clusters first, in some cases analogously to the hierarchical branching of tree diagrams (*14, 18*), although sample sizes and within-group diversity levels also affect splitting order (*18*). One strategy for analysis is to apply *structure* for many values of K (the number of clusters) and to select the K that maximizes the posterior probability of the data (*14*). For very complex datasets that include many groups, this criterion is difficult to apply: the algorithm may converge to numerous distinct clustering schemes for a given value of K, so that estimated probabilities differ across runs (*18*). Consequently, as multiple clustering solutions appeared for $K \geq 7$ (many similarity coefficients below 0.85, with different groups comprising the "additional" clusters in different runs), we used small K to analyze population structure in the worldwide sample and we subdivided the sample for further analysis.

Similarity coefficients: The coefficient $C(\mathbf{Q}_1, \mathbf{Q}_2) = 1 - (min_P ||\mathbf{Q}_1 - P(\mathbf{Q}_2)||_F)/||\mathbf{Q}_1 - \mathbf{1/K}||_F$ quantifies the similarity of results for an ordered pair of *structure* runs with the same number of assumed clusters K. The $I \times K$ row-stochastic matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$, where I is the number of individuals, represent estimated membership fractions for the K clusters in the two runs. P is a permutation of the columns of a matrix, the minimum is taken over the $K!$ permutations of the clusters (columns) of $\mathbf{Q}_2$, $|| \cdot ||_F$ is the Frobenius matrix norm (*33*), and $\mathbf{1/K}$ is the $I \times K$ matrix with all entries equal to $1/K$. If the two runs had different numbers of individuals (as when runs with reduced samples were compared to runs with the full data), only rows of the larger matrix that corresponded to individuals represented in the smaller matrix were used.

Values of C for a pair of runs roughly correspond to the following descriptions: 0.85-1.0, nearly all individuals have nearly identical membership coefficients in both runs; 0.4-0.85, most individuals have similar membership coefficients, but the other individuals may have very different placements; 0.1-0.4, some of the inferred clusters consist of the same sets of individuals in both runs, but the other clusters differ greatly across runs; <0.1, inferred population structures have few similarities. $C < 0$ is possible, though this was almost never observed.

Similarity coefficients for reduced data: For America, Oceania, Africa, Middle East, and the worldwide sample, the median similarity coefficients $C(\mathbf{Q}_{377}, \mathbf{Q}_L)$ were computed for 100 comparisons of 10 runs using all loci with runs using each of 10 sets of L random loci. For 377 loci the 90 comparisons of the 10 different full-data runs were used, and $C = 0$ was assumed with no data. Median similarity coefficients were computed between runs with the full data and runs with both the number of loci and the number of individuals reduced. For reduced samples, half of the individuals in each population were chosen randomly, rounding up if appropriate.

Supporting text

Atypical individuals: Biaka #980 and Japanese #770 were inferred to be particularly atypical for their populations. Using the *structure* migration model with the worldwide sample, $K = 6$ and a migration prior of 0.0001 for all individuals (*14*), both individuals had posterior probability 1 of having had contaminated or mislabeled samples, or of having been migrants (#980 from Eurasia and #770 from America).

Splitting order: The order in which American populations split, observed in all runs, was as follows (not shown). At $K = 2$, one cluster contained Karitiana and Surui from South America, two isolated groups with low expected heterozygosity (0.571 and 0.501, respectively, compared to the population average of 0.727). At $K = 3$, Karitiana and Surui split into separate clusters, and at $K = 4$, Colombians comprised the new cluster. With $K = 2$ and the African sample, Biaka separated from other populations, and at $K = 3$, a joint Mbuti-San group separated (not shown). The Middle East was the only region for which the number of inferred clusters was consistently larger than the number of predefined populations. For four of the clusters, membership was largely limited to a single population; some individuals from each group, especially Bedouin, had large membership coefficients in the fifth cluster. At $K = 2$ and $K = 3$ most populations of the Middle East had considerable membership in all clusters (not shown). At $K = 4$, three clusters were largely restricted to Bedouin, Druze, and Mozabite, respectively; the fourth cluster had partial membership from all populations.

For the worldwide sample, the observation that Africans (or a subset thereof) did not separate from all other populations at $K = 2$ might reflect their small sample size compared to Eurasians, and does not argue against an ancient African divergence. Runs of *structure* at $K = 2$ using subsamples of the data with comparable sample sizes from Africa, Eurasia, East Asia, and America placed Africans and Americans as anchors of the two clusters, with all other individuals exhibiting significant membership coefficients in both clusters (not shown); at $K = 3$ Africans and Americans each formed separate clusters.

Multiple clustering solutions: For samples in which clustering solutions were somewhat uncertain (East Asia, Eurasia, Central/South Asia, Europe), several lines of evidence suggest that the inferred clusters do not simply reflect random inference of population structure where there was no genuine signal. First, with unstructured data and K clusters, *structure* typically assigns membership coefficients of approximately $1/K$ for each individual and each cluster (see the *structure* manual at http://pritch.bsd.uchicago.edu). By contrast, with one exception (Europe, $K = 3$), the observed distribution of membership coefficients across clusters was highly asymmetric: individuals usually had one or two large membership coefficients, with the others small. Additionally, for all regions except Europe, runs with $K > 1$ almost always produced higher posterior probabilities than those with $K = 1$. Even for Europe, the runs of highest probability had $K > 1$.

Heterozygosity: As has been previously observed, Africa was the most variable region (*7, 12, 29*), with average within-population heterozygosity equal to 0.774 (Supplementary Table 3). The high African diversity was reflected more dramatically in the geographic distribution of alleles: more than half of region-specific alleles were unique to Africa (Supplementary Figure 1). Also, the populations with the most private alleles were African: Biaka, Mbuti and San. This observation is particularly interesting in light of the small number of San in the sample (7 individuals). A relatively large number of alleles was found in all populations except San (63 alleles). This observation might result from the small sample size; however, Tuscans, represented by 8 individuals, were missing far fewer of these otherwise universal alleles (14 alleles). The distinctiveness of Biaka, Mbuti, and San is consistent with their putatively ancient divergence from other populations (*34*).

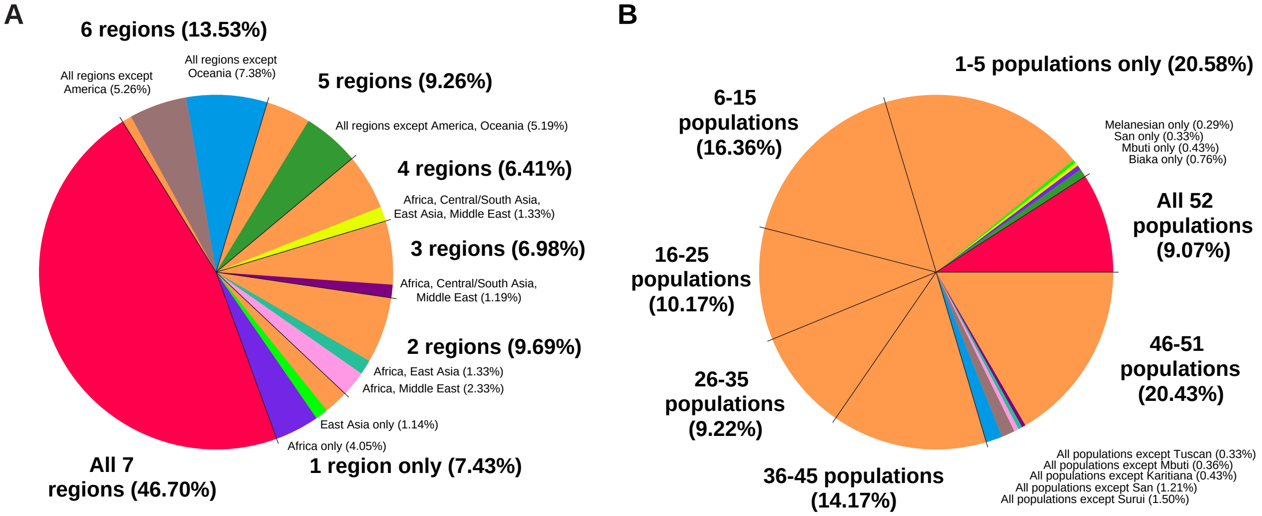
Supporting references

31. Lewis, P. O. and Zaykin, D. GDA (Genetic Data Analysis): Computer Program for the Analysis of Allelic Data (version 1.0 d16c). http://lewis.eeb.uconn.edu/lewishome/software.html, 2001.

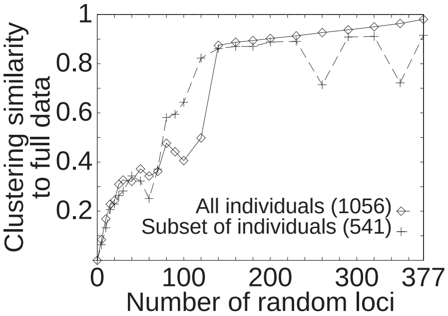
32. B. S. Weir, *Genetic Data Analysis II* (Sinauer, Sunderland, MA, 1996).

33. G. H. Golub, C. F. Van Loan, *Matrix Computations* (Johns Hopkins University Press, Baltimore, 1996).

34. F. Cruciani et al., *Am. J. Hum. Genet.* **70**, 1197 (2002).



Supplementary Figure 1: Classification of 4199 non-singleton alleles. (A) Each allele was classified based on its presence or absence in seven predefined regions. Of $2^7 - 1 = 127$ possible presence/absence categories, 112 types of alleles were observed. Extended radii (black) divide the circle based on the number of regions in which alleles were found. Categories with the largest numbers of alleles are shown explicitly, and remaining categories are grouped (orange). 312 alleles were private to one region, distributed as follows: 170 (Africa), 48 (East Asia), 29 (Middle East), 29 (Central/South Asia), 15 (Oceania), 12 (America), and 9 (Europe). Of 2864 alleles with 20 or more copies in the sample (that is, alleles with frequency of at least $\sim 1\%$), 68.2% were present in all regions. (B) Each allele was classified based on presence or absence in each of the 52 populations. Of $2^{52} - 1$ ($\sim 10^{15}$) possible categories, 3264 were observed. Extended radii (black) divide the circle based on the number of *populations* in which alleles were found. Categories with the largest numbers of alleles are shown explicitly, and remaining categories (orange) are grouped. 146 alleles were private to one population. These alleles were distributed into regions as follows: 76 (Africa), 18 (Middle East), 15 (Central/South Asia), 14 (Oceania), 11 (East Asia), 9 (America), 3 (Europe).



Supplementary Figure 2: Similarity of *structure* runs for reduced samples to those based on the full worldwide sample. Analogously to corresponding plots in Figure 2, median similarity coefficients to runs with the full sample are displayed for runs with only the number of loci reduced, and for runs with both the number of loci and the number of individuals reduced.

Supplementary Table 1: Genetic distances between regional groups. The coancestry or F_{st} distance (*32*, p. 194) was estimated using GDA (*31*).

	Africa	**Europe**	**Middle East**	**Central/ South Asia**	**East Asia**	**Oceania**
Europe	0.040					
Middle East	0.033	0.005				
Central/South Asia	0.037	0.008	0.008			
East Asia	0.054	0.038	0.038	0.026		
Oceania	0.068	0.061	0.059	0.049	0.047	
America	0.101	0.079	0.081	0.068	0.060	0.102

Supplementary Table 2: Membership coefficients for the $K = 6$ clustering shown in Figure 1, averaged across individuals. Average membership coefficients across individuals are also shown for geographic regions.

Population	Orange	Blue	Yellow	Pink	Green	Purple
Bantu (Kenya)	0.89	0.09	0.01			0.01
Mandenka	0.97	0.01	0.01	0.01	0.01	
Yoruba	0.96	0.02	0.01	0.01		
San	0.98	0.01			0.01	
Mbuti Pygmy	0.99	0.01				
Biaka Pygmy	0.97	0.02		0.01		
Africa	0.96	0.03		0.01		
Orcadian		0.98	0.01			0.01
Adygei		0.94	0.02	0.02	0.01	0.01
Russian		0.93	0.01	0.03	0.01	0.01
Basque		0.98	0.01	0.01		0.01
French		0.97	0.01	0.01		0.01
Italian		0.98				
Sardinian		0.99				
Tuscan		0.99	0.01			
Europe		0.97	0.01	0.01		0.01
Mozabite	0.23	0.76				
Bedouin	0.06	0.93		0.01	0.01	
Druze		0.98	0.01	0.01		
Palestinian	0.02	0.95	0.01	0.01	0.01	
Middle East	0.06	0.92	0.01	0.01		
Balochi	0.02	0.90	0.04	0.04	0.01	
Brahui	0.02	0.90	0.03	0.04	0.01	0.01
Makrani	0.05	0.84	0.05	0.04	0.02	
Sindhi	0.03	0.81	0.07	0.06	0.02	0.01
Pathan		0.79	0.08	0.09	0.02	0.02
Burusho		0.69	0.10	0.17	0.02	0.02
Hazara		0.52	0.01	0.45	0.01	0.01
Uygur		0.42	0.04	0.53	0.01	0.01
Kalash		0.16	0.83			
Central/South Asia	0.01	0.69	0.15	0.13	0.01	0.01
Han				0.98	0.01	0.01
Han (N. China)		0.02	0.01	0.96	0.01	0.01
Dai		0.01		0.97	0.02	
Daur		0.02	0.01	0.96	0.01	0.01
Hezhen				0.98	0.01	0.01
Lahu			0.01	0.97	0.01	
Miao				0.98	0.01	
Oroqen		0.02	0.01	0.95		0.02
She				0.99		
Tujia		0.01		0.98		0.01
Tu	0.01	0.03	0.01	0.93	0.01	0.01
Xibo		0.05	0.01	0.92		0.02
Yi	0.01	0.01		0.97	0.01	0.01
Mongola		0.03	0.01	0.93	0.01	0.02
Naxi		0.01	0.01	0.97	0.01	0.01
Cambodian	0.01	0.06	0.02	0.88	0.03	0.01
Japanese		0.01		0.94	0.02	0.03
Yakut		0.10	0.02	0.87		0.01
East Asia		0.02	0.01	0.95	0.01	0.01
Melanesian				0.03	0.96	
Papuan	0.01	0.05	0.01	0.12	0.80	0.01
Oceania	0.01	0.02		0.07	0.89	
Karitiana				0.01		0.99
Surui						1.00
Colombian		0.02	0.01	0.12		0.85
Maya	0.01	0.12	0.02	0.18	0.01	0.66
Pima		0.01		0.07	0.01	0.91
America		0.03	0.01	0.07		0.88

Supplementary Table 3: Population summary statistics. Expected heterozygosity was estimated for each locus and was averaged across loci. The unbiased estimator $[2n/(2n-1)][1-\sum_{i=1}^{k}\hat{p}_i^2]$, was used, where n is the number of individuals, k is the number of distinct alleles, and $\hat{p}_i$ is the relative frequency of allele i in the sample. The number of observed alleles was averaged across loci. The statistics were also calculated regionally, by grouping all populations from each region.

Population	Sample size	Heterozygosity	Number of alleles
Bantu (Kenya)	12	0.782	6.38
Mandenka	24	0.776	7.46
Yoruba	25	0.780	7.50
San	7	0.762	5.09
Mbuti Pygmy	15	0.770	6.56
Biaka Pygmy	36	0.775	7.72
Africa (average across populations)	19.8	0.774	6.79
Africa (treated as one region)	119	0.792	10.15
Orcadian	16	0.747	5.93
Adygei	17	0.755	6.28
Russian	25	0.754	6.62
Basque	24	0.746	6.45
French	29	0.753	6.88
Italian	14	0.750	5.88
Sardinian	28	0.749	6.70
Tuscan	8	0.754	5.13
Europe (average across populations)	20.1	0.751	6.23
Europe (treated as one region)	161	0.753	8.80
Mozabite	30	0.762	7.16
Bedouin	49	0.757	7.64
Druze	48	0.748	7.26
Palestinian	51	0.758	7.65
Middle East (average across populations)	44.5	0.756	7.43
Middle East (treated as one region)	178	0.761	9.35
Balochi	25	0.758	6.65
Brahui	25	0.754	6.71
Makrani	25	0.763	6.97
Sindhi	25	0.759	6.81
Pathan	25	0.756	6.79
Burusho	25	0.751	6.69
Hazara	25	0.752	6.78
Uygur	10	0.753	5.55
Kalash	25	0.721	5.86
Central/South Asia (average across populations)	23.3	0.752	6.53
Central/South Asia (treated as one region)	210	0.759	9.34
Han	35	0.724	6.80
Han (N. China)	10	0.730	5.28
Dai	10	0.722	5.14
Daur	10	0.731	5.22
Hezhen	10	0.718	4.95
Lahu	10	0.699	4.80
Miao	10	0.717	5.16
Oroqen	10	0.723	5.10
She	9	0.709	4.88
Tujia	10	0.718	5.17
Tu	10	0.728	5.30
Xibo	9	0.735	5.14
Yi	10	0.732	5.26
Mongola	10	0.730	5.30
Naxi	10	0.713	5.07
Cambodian	11	0.732	5.48
Japanese	32	0.721	6.68
Yakut	25	0.726	6.25
East Asia (average across populations)	13.4	0.723	5.39
East Asia (treated as one region)	241	0.730	9.26
Melanesian	22	0.668	5.17
Papuan	17	0.698	5.49
Oceania (average across populations)	19.5	0.683	5.33
Oceania (treated as one region)	39	0.695	6.46
Karitiana	24	0.571	4.02
Surui	21	0.501	3.28
Colombian	13	0.615	4.17
Maya	25	0.689	5.90
Pima	25	0.617	4.53
America (average across populations)	21.6	0.599	4.38
America (treated as one region)	108	0.664	6.80
World (average across populations)	20.3	0.727	5.94
World (treated as one region)	1056	0.771	12.42

Supplementary Table 4. Summary statistics for loci. To estimate expected heterozygosity, $[2n/(2n-1)][1-\sum_{i=1}^{k}\hat{p}_i^2]$, was used, where n is the number of individuals, k is the number of distinct alleles, and $\hat{p}_i$ is the relative frequency of allele i in the sample. All autosomal loci in Marshfield Screening Set #10 are shown, except D11S1985 (also known as GGAA5C04), which was not genotyped. For loci that begin with "NA," the alternate names should be used. Loci are sorted by heterozygosity. The average heterozygosity was 0.771 (standard deviation 0.065) and the average number of alleles was 12.42 (standard deviation 4.11). Heterozygosity did not vary significantly across chromosomes ($P = 0.39$, Kruskal-Wallis test), nor did number of alleles ($P = 0.16$).

Locus name	Alternate name	Heterozygosity	Number of alleles	Chromosome
D3S2427	GATA22F11	0.907	29	3
D21S2055	GATA188F04	0.907	24	21
D22S683	GATA11B12	0.899	32	22
D11S1986	GGAA7G08	0.897	23	11
D1S3721	GATA129H04	0.895	17	1
D15S822	GATA88H02	0.895	19	15
D2S1334	GATA4D07	0.894	20	2
D20S159	UT1307	0.894	20	20
D11S2000	GATA28D01	0.893	25	11
D3S2387	GATA22G12	0.892	27	3
D2S1788	GATA86E02	0.888	15	2
D7S1804	GATA43C11	0.888	18	7
D4S2632	GATA72G09	0.886	21	4
D16S3401	16PTEL06	0.881	20	16
D3S1746	GATA8F01	0.873	26	3
NA-D12S-1	GATA49D12	0.872	22	12
D21S1411	UT1355	0.872	16	21
D13S285	AFM309VA9	0.869	18	13
D7S3046	GATA118G10	0.869	19	7
D20S851	AFMA218YB5	0.868	15	20
D11S1984	GGAA17G05	0.866	18	11
D2S1399	GGAA20G04	0.865	12	2
D9S1838	AFMB303ZG9	0.865	16	9
D1S534	GATA12A07	0.864	18	1
D8S1132	GATA26E03	0.863	14	8
D15S659	GATA63A03	0.860	16	15
D1S1612	GGAA3A07	0.859	12	1
NA-D18S-2	ATA82B02	0.857	12	18
NA-D14S-1	GATA193A07	0.857	14	14
D16S422	AFM249XC5	0.854	21	16
D15S128	AFM273YF9	0.853	14	15
D1S1679	GGAA5F09	0.853	19	1
D1S518	GATA7C01	0.852	11	1
D15S652	ATA24A08	0.851	15	15
D12S297	UT5029	0.851	15	12
D7S2204	GATA73D10	0.850	15	7
D18S1364	GATA7E12	0.849	13	18
D14S587	GGAA10C09	0.849	12	14
D7S3058	GATA30D09	0.849	11	7
NA-D22S-1	GATA198B05	0.848	13	22
D3S1311	AFM254VE1	0.847	16	3
D20S478	GATA42A03	0.846	13	20
D4S2623	GATA62A12	0.846	14	4
D12S269	MFD259	0.845	14	12
D9S2157	ATA59H06	0.845	13	9
D6S1056	GATA68H04	0.844	12	6
D2S410	GATA4E11	0.844	16	2
D15S642	GATA27A03	0.843	16	15
D2S1356	ATA4F03	0.843	10	2
D15S643	GATA50G06	0.842	19	15
D6S305	AFM242ZG5	0.840	18	6
D17S1290	GATA49C09	0.839	18	17
D14S1007	AFMB002ZF1	0.838	15	14
D5S2505	GATA84E11	0.838	15	5
NA-D6S-1	GATA184A08	0.838	14	6
D18S542	GATA11A06	0.838	20	18
D2S1353	ATA27H09	0.837	11	2
D6S2439	GATA163B10	0.837	13	6
D20S451	UT254	0.837	19	20
D9S1118	GATA71E08	0.836	24	9
D3S1560	AFM217XD6	0.836	18	3
D10S677	GGAA2F11	0.836	12	10
D7S1808	GGAA3F06	0.835	12	7
D5S1470	GATA7C06	0.834	14	5
D14S608	GATA43H01	0.834	11	14
D21S2052	GATA129D11	0.833	12	21
D12S1042	ATA27A06	0.832	9	12
D17S2196	GATA185H04	0.832	10	17
D5S1505	GATA62A04	0.830	15	5
D3S4545	GATA164B08	0.830	28	3

Locus name	Alternate name	Heterozygosity	Number of alleles	Chromosome
D3S1262	AFM059XA9	0.829	14	3
D1S1609	GATA50F11	0.829	15	1
D19S433	GGAA2A03	0.829	17	19
D5S211	MFD154	0.829	14	5
D8S1179	GATA7G07	0.828	11	8
D18S1370	ATA45G06	0.827	10	18
D7S3070	GATA189C06	0.827	14	7
D5S1480	ATA23A10	0.826	12	5
D3S3630	AFMB296ZF5	0.825	18	3
D3S3045	GATA84B12	0.825	11	3
NA-D16S-1	GATA138C05	0.823	20	16
D16S2621	GATA71F09	0.823	10	16
D11S1999	GATA23F06	0.821	10	11
D3S1744	GATA3C02	0.821	11	3
D22S1169	AFMB337ZH9	0.820	11	22
D4S2366	GATA22G05	0.820	8	4
D7S821	GATA5D08	0.820	11	7
D7S3061	GGAA6D03	0.819	13	7
D19S246	MFD232	0.819	15	19
D5S2500	GATA67D03	0.818	12	5
NA-D4S-1	GATA70E01	0.818	17	4
NA-D18S-1	GATA178F11	0.818	12	18
D9S301	GATA7D12	0.817	14	9
D7S1824	GATA32C12	0.817	12	7
D20S477	GATA29F06	0.817	11	20
D9S938	GGAA22E01	0.816	12	9
D3S2432	GATA27C08	0.815	13	3
D12S1064	GATA63D12	0.815	14	12
D8S373	UT721	0.815	11	8
D9S925	GATA27A11	0.814	17	9
D11S1981	GATA48E02	0.813	13	11
D18S1357	ATA7D07	0.813	11	18
D16S403	AFM049XD2	0.813	14	16
D10S1221	ATA21A03	0.812	13	10
D20S481	GATA47F05	0.812	11	20
D8S1477	GGAA20C10	0.812	14	8
D13S796	GATA51B02	0.812	12	13
D16S753	GGAA3G05	0.811	12	16
D5S816	GATA2H09	0.811	10	5
D13S1493	GGAA29H03	0.809	12	13
D2S1360	GATA11H10	0.809	15	2
D13S317	GATA7G10	0.809	9	13
D1S2134	GATA72H07	0.808	14	1
D12S2070	ATA25F09	0.808	8	12
D14S1426	GATA136B01	0.808	19	14
D2S2944	GATA30E06	0.807	9	2
D14S599	ATA29G03	0.807	10	14
D8S1128	GATA21C12	0.807	11	8
D1S1660	GATA48B01	0.806	9	1
D4S1627	GATA7D01	0.806	8	4
NA-D1S-1	ATA79C10	0.806	12	1
F13A1-D6S	SE30	0.806	17	6
D4S2394	ATA26B08	0.805	11	4
D11S2362	ATA33B03	0.805	13	11
D7S817	GATA13G11	0.804	11	7
D11S1993	ATA1B07	0.804	12	11
D9S1871	AFM345TA9	0.803	23	9
D14S617	GGAA21G11	0.803	11	14
D5S1462	GATA3H06	0.802	12	5
D7S2477	AFMB035XB9	0.802	18	7
D14S592	ATA19H08	0.802	11	14
D7S559	MFD265	0.801	20	7
D9S930	GATA48D07	0.801	9	9
D2S2952	GATA116B01	0.801	17	2
D2S1384	GATA52A04	0.800	15	2
D3S1768	GATA8B05	0.800	10	3
D4S408	AFM165XC11	0.800	13	4
D6S1031	ATA28B11	0.800	11	6
D6S1053	GATA64D02	0.800	10	6
D12S1045	ATA29A06	0.799	11	12
D2S1363	GATA23D03	0.799	10	2
D4S1647	GATA2F11	0.798	10	4
NA-D8S-1	GATA151F02	0.798	17	8
D12S1294	GATA73H09	0.797	13	12
D16S3396	ATA55A11	0.797	12	16
D10S2470	GATA115E01	0.796	10	10
D9S922	GATA21F05	0.795	9	9
D1S3669	GATA29A05	0.795	12	1
D16S516	AFM350VD1	0.795	12	16
D1S1589	ATA4E02	0.794	11	1
D1S235	AFM203YG9	0.794	19	1
D10S1430	GATA84C01	0.794	11	10
D4S2431	GGAA19H07	0.794	22	4
D1S549	GATA4H09	0.793	11	1
D20S201	GATA8B01	0.792	14	8

Locus name	Alternate name	Heterozygosity	Number of alleles	Chromosome
D11S1304	UT2095	0.792	15	11
D11S2363	GATA12F04	0.791	19	11
NA-D12S-2	PAH	0.791	10	12
D12S1301	GATA91H06	0.791	9	12
D8S1110	GATA8G10	0.790	14	8
D7S820	GATA3F01	0.790	16	7
D18S535	GATA13	0.790	10	18
D3S2398	GATA6G12	0.790	9	3
D16S539	GATA11C06	0.790	10	16
D2S2986	2QTEL47	0.790	12	2
D9S1121	GATA87E02	0.790	15	9
D11S2002	GATA30G01	0.790	10	11
D8S2324	GATA14E09	0.789	9	8
D22S689	GATA21F03	0.788	14	22
D8S1113	GGAA8G07	0.788	10	8
D5S1501	GATA52A12	0.787	17	5
D22S345	MFD313	0.787	17	22
D5S1725	GATA89G08	0.787	11	5
D17S2193	ATA43A10Z	0.786	20	17
D21S1437	GGAA3C07	0.785	12	21
D1S3462	ATA29C07	0.785	11	1
D13S793	GATA43H03	0.785	8	13
D6S2436	GATA165G02	0.783	11	6
D4S2397	ATA27C07	0.783	9	4
D14S588	GGAA4A12	0.782	10	14
D19S254	MFD238	0.782	14	19
D10S1248	GGAA23C05	0.782	14	10
D20S480	GATA45B10	0.782	9	20
D5S1457	GATA21D04	0.781	12	5
D1S468	AFM280WE5	0.781	13	1
D17S784	AFM044XG3	0.780	14	17
D4S403	AFM157XG3	0.780	13	4
D5S1456	GATA11A11	0.779	14	5
D5S820	GATA6E05	0.778	13	5
D2S1328	GATA27A12	0.778	9	2
D6S1027	ATA22G07	0.778	11	6
D19S714	GATA66B04	0.778	10	19
D2S1790	GATA88G05	0.777	18	2
D12S2078	GATA32F05	0.777	10	12
D9S934	GATA64G07	0.776	12	9
D11S4463	GATA117D01	0.775	10	11
D2S441	GATA8F03	0.775	15	2
D5S2849	GATA145D10	0.773	10	5
D9S1825	AFMB029XG1	0.773	17	9
D19S559	UT7544	0.773	9	19
D2S434	GATA4G12	0.772	10	2
NA-D13S-1	ATA5A09	0.771	10	13
D6S474	GATA31	0.771	8	6
D22S686	GGAA10F06	0.771	12	22
NA-D10S-1	GATA121A08	0.771	10	10
D10S1230	ATA29C03	0.770	11	10
D5S408	AFM164XB8	0.770	15	5
D3S4529	GATA128C02	0.770	6	3
D13S800	GATA64F08	0.769	12	13
D10S1208	ATA5A04	0.768	12	10
D3S1763	GATA3H01	0.768	10	3
D1S3720	ATA47D07	0.767	9	1
D10S1225	ATA24F10	0.766	10	10
D6S1277	GATA81B01	0.766	11	6
D3S1764	GATA4A10	0.765	14	3
D11S969	AFM205VF10	0.765	14	11
D19S591	GATA44F10	0.763	9	19
D7S3051	GATA137H02	0.763	10	7
NA-D15S-1	GATA50C03	0.761	13	15
D10S1423	GATA70E11	0.761	10	10
D2S1776	GATA71D01	0.759	10	2
D22S1045	ATA37D06	0.758	10	22
D4S3248	GATA28F03	0.758	10	4
D16S748	ATA3A07	0.758	12	16
D11S4464	GATA64D03	0.758	11	11
D4S2368	GATA27G03	0.757	8	4
D2S427	GATA12H10	0.757	10	2
D12S1300	GATA85A04	0.757	10	12
D17S1294	GGAA9D03	0.755	13	17
D21S1432	GATA11C12	0.754	15	21
D19S589	GATA29B01	0.754	9	19
D6S1040	GATA23F08	0.754	8	6
D11S1392	GATA6B09	0.752	11	11
D4S3243	GATA10G07	0.751	11	4
NA-D1S-4	ATA42G12	0.750	9	1
D7S2846	GATA31A10	0.750	8	7
D21S1446	GATA70B08	0.749	14	21
D5S2845	GATA134B03	0.749	9	5
D3S2460	GATA68F07	0.748	9	3
D4S2367	GATA24H01	0.748	10	4

Locus name	Alternate name	Heterozygosity	Number of alleles	Chromosome
D6S1017	GGAT3H10	0.748	9	6
D6S1009	GATA32B03	0.748	13	6
D8S261	AFM123XG5	0.747	17	8
D18S1390	18QTEL11	0.747	15	18
D13S1807	GATA11C08	0.747	9	13
D12S1052	GATA26D02	0.747	8	12
D16S2616	ATA41E04	0.746	11	16
D2S1780	GATA72G11	0.746	13	2
NA-D1S-3	GATA133A08	0.745	12	1
NA-D8S-2	GAAT1A4	0.745	9	8
D4S1625	GATA107	0.745	10	4
D18S851	GATA6D09	0.745	12	18
D8S1136	GATA41A01	0.745	11	8
NA-D5S-1	ATA52D02	0.744	27	5
D3S3038	GATA73D01	0.743	14	3
D15S1515	GATA197B10	0.743	9	15
D16S2624	GATA81D12	0.742	8	16
D2S2972	GATA176C01	0.741	14	2
D13S895	GGAA22G01	0.740	11	13
D11S1998	GATA23E06	0.740	9	11
D10S2327	GGAT1A4	0.740	9	10
D17S974	GATA8C04	0.739	8	17
D4S2361	ATA2A03	0.739	16	4
D4S1629	GATA8A05	0.737	8	4
D10S1435	GATA88F09	0.737	14	10
NA-D17S-1	ATA78D02	0.736	10	17
D18S858	ATA23G05	0.736	10	18
D2S1394	GATA69E12	0.735	9	2
D8S560	AFMA127YE5	0.734	20	8
D16S3253	GATA22F09	0.734	12	16
D6S2410	GATA11E02	0.734	10	6
D5S2488	ATA20G07	0.734	12	5
D7S1818	GATA24D12	0.733	9	7
D12S395	GATA4H01	0.732	12	12
D1S1677	GGAA22G10	0.731	11	1
D12S372	GATA4H03	0.731	10	12
D1S1596	GATA26G09	0.731	8	1
D8S503	AFM193XH4	0.730	13	8
D10S1432	GATA87G01	0.730	11	10
D12S1638	AFMB002VD5	0.730	11	12
D10S189	AFM063XF4	0.730	6	10
D5S2501	GATA68A03	0.730	10	5
D10S1222	ATA22D02	0.730	11	10
D12S373	GATA6C01	0.729	9	12
D14S742	GATA74E02	0.729	11	14
D20S164	UT1772	0.729	10	20
D7S1799	GATA23F05	0.729	12	7
D9S1120	GATA81C04	0.729	14	9
NA-D7S-1	GATA104	0.728	16	7
D7S3056	GATA24F03	0.727	13	7
D9S1779	AFM026TG9	0.727	19	9
D20S1143	GATA129B03	0.727	9	20
D18S843	ACT1A01	0.726	8	18
D4S3360	4PTEL04	0.726	11	4
D4S1644	GATA11E09	0.725	13	4
D10S1426	GATA73E11	0.724	10	10
D9S2169	GATA62F03	0.722	10	9
D1S551	GATA6A05	0.722	11	1
NA-D9S-1	GATA187D09	0.722	9	9
D19S1034	GATA21G05	0.722	17	19
D7S1802	GATA41G07	0.721	12	7
D18S877	GATA64H04	0.721	9	18
D3S1766	GATA6F06	0.721	10	3
D15S816	GATA73F01	0.720	10	15
D9S910	ATA18A07	0.720	10	9
D2S1391	GATA65C03	0.720	14	2
D1S1597	GATA27E01	0.719	9	1
D15S818	GATA85D02	0.719	9	15
D15S165	AFM248VC5	0.717	20	15
D19S586	GATA23B01	0.716	10	19
D21S1440	ATA27F01	0.716	12	21
D8S592	GATA6B02	0.713	9	8
D3S3039	GATA7F05	0.713	9	3
D2S1352	ATA27D04	0.712	9	2
D9S1122	GATA89A11	0.712	8	9
D15S655	ATA28G05	0.711	10	15
D15S1507	GATA151F03	0.711	8	15
D3S4523	ATA34G06	0.710	11	3
NA-D11S-1	ATA34E08	0.709	11	11
D10S1239	GATA64A09	0.708	10	10
D7S3047	GATA119B03	0.707	9	7
D14S1434	GATA168F06	0.707	8	14
D1S1627	ATA25E07	0.706	7	1
D1S1665	GATA61A06	0.705	17	1
D20S482	GATA51D03	0.704	9	20

Locus name	Alternate name	Heterozygosity	Number of alleles	Chromosome
NA-D1S-5	GATA124F08	0.704	8	1
D8S1048	UT7129	0.703	10	8
D10S1412	ATA31G11	0.700	9	10
D10S1425	GATA71C09	0.699	8	10
D19S245	MFD235	0.699	6	19
D14S1280	GATA31B09	0.699	15	14
D6S1051	GATA61E03	0.698	8	6
TPO-D2S	SRA	0.698	8	2
D14S606	GATA30A03	0.698	9	14
D18S1371	GATA177C03	0.696	8	18
D11S2365	GATA63F09	0.693	10	11
D3S2403	GGAA4B09	0.691	20	3
D4S1652	GATA5B02	0.691	8	4
D5S817	GATA3E10	0.689	10	5
D6S1959	GATA29A01	0.687	7	6
D1S1728	GATA109	0.687	9	1
D17S1301	GATA28D11	0.686	10	17
D11S2371	GATA90D07	0.686	10	11
D16S764	GATA42E11	0.686	9	16
D1S1653	GATA43A04	0.686	11	1
D2S405	GATA8F07	0.684	8	2
D3S2409	ATA10H11	0.684	10	3
D13S787	GATA23C03	0.683	12	13
D16S769	GATA71H05	0.683	8	16
D11S2006	GATA46A12	0.676	9	11
D17S1299	GATA25A04	0.674	11	17
D2S1400	GGAA20G10	0.673	15	2
D8S1108	GATA50D10	0.672	9	8
D8S262	AFM127XH2	0.669	10	8
D17S1308	GTAT1A05	0.666	7	17
D11S4459	ATA9B04	0.664	8	11
D3S2418	ATA22E01	0.662	12	3
D6S942	UT654	0.661	11	6
D13S779	ATA26D07	0.661	10	13
D20S103	AFM077XD3	0.661	14	20
D13S894	GATA86H01	0.660	10	13
D22S532	UT7136	0.660	9	22
NA-D1S-2	GATA124C08	0.657	9	1
D4S2417	GATA42H02	0.655	8	4
D1S1594	GATA22D12	0.654	10	1
D6S1021	ATA11D10	0.643	11	6
D1S2682	AFMA272XC9	0.632	12	1
NA-D10S-2	ATA103C06	0.630	11	10
D2S2968	GATA178G09	0.626	11	2
D17S2195	ATA58A02	0.626	17	17
D10S212	AFM198ZB4	0.613	10	10
D3S3644	AFMB318YF1	0.605	9	3
D6S1006	ATC4D09	0.585	4	6
D17S2180	ATC6A06	0.570	9	17
D18S1376	GATA185C06	0.564	8	18
D17S1298	GAAT2C03	0.557	5	17
D6S2522	6QTEL54	0.496	5	6