European Journal of Human Genetics (2003) 11, 39–49
© 2003 Nature Publishing Group All rights reserved 1018–4813/03 $25.00
www.nature.com/ejhg

ARTICLE

# Global genetic variation at nine short tandem repeat loci and implications on forensic genetics

Guangyun Sun[1], Stephen T McGarvey[2], Riad Bayoumi[3], Connie J Mulligan[4], Ramiro Barrantes[5], Salmo Raskin[6], Yixi Zhong[7], Joshua Akey[1], Ranajit Chakraborty[1] and Ranjan Deka*[,1]

[1]Center for Genome Information, Department of Environmental Health, University of Cincinnati, Cincinnati, Ohio, USA; [2]International Health Institute and Department of Community Health, Brown University, Providence, Rhode Island, USA; [3]Department of Biochemistry, Sultan Qaboos University, Muscat, Oman; [4]Department of Anthropology, University of Florida, Gainesville, Florida, USA; [5]Instituto de Investigaciones en Salud, Universidad de Costa Rica, Costa Rica; [6]Laboratorio Genetika, Curitiba, Brazil; [7]Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Texas, USA

**We have studied genetic variation at nine autosomal short tandem repeat loci in 20 globally distributed human populations defined by geographic and ethnic origins, viz., African, Caucasian, Asian, Native American and Oceanic. The purpose of this study is to evaluate the utility and applicability of these nine loci in forensic analysis in worldwide populations. The levels of genetic variation measured by number of alleles, allele size variance and heterozygosity are high in all populations irrespective of their effective sizes. Single- as well as multi-locus genotype frequencies are in conformity with the assumptions of Hardy-Weinberg equilibrium. Further, alleles across the entire set of nine loci are mutually independent in all populations. Gene diversity analysis shows that pooling of population data by major geographic groupings does not introduce substructure effects beyond the levels recommended by the National Research Council, validating the establishment of population databases based on major geographic and ethnic groupings. A network tree based on genetic distances further supports this assertion, in which populations of common ancestry cluster together. With respect to the power of discrimination and exclusion probabilities, even the relatively reduced levels of genetic variation at these nine STR loci in smaller and isolated populations provide an exclusionary power over 99%. However, in paternity testing with unknown genotype of the mother, the power of exclusion could fall below 80% in some isolated populations, and in such cases use of additional loci supplementing the battery of the nine loci is recommended.**

*European Journal of Human Genetics* (2003) **11,** 39–49. doi:10.1038/sj.ejhg.5200902

**Keywords:** forensic genetics; population genetics; STR database; forensic markers; parentage testing; Hardy–Weinberg equilibrium

## Introduction

The remarkable progress made in DNA technology in the past decade has had an enormous impact on several disciplines, including forensic science. Identification of thousands of genetic markers, particularly the short tandem repeat (STR) loci, distributed throughout the human genome, and their analysis using polymerase chain reaction (PCR) based techniques, tremendously augmented the efficiency in individual identification and determination of genetic relationships among individuals. Based on population genetic characteristics desired in forensic analysis, such as adherence to the expectations of Hardy–Weinberg equilibrium (HWE) and independence of alleles across loci,

*Correspondence: R Deka, Department of Environmental Health, University of Cincinnati, 3223 Eden Avenue, Cincinnati, OH 45267-0056, USA. Tel: +1 513 558 5989; Fax: +1 513 558 4397;
E-mail: ranjan.deka@uc.edu

Received 10 June 2002; revised 26 August 2002; accepted 27 August 2002

as well as ease of laboratory typing, a set of 13 STR loci (viz., D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539, CSF1P0, TPOX, TH01) have been established as the core genetic markers for use in DNA forensic analysis and parentage testing.[1,2] These developments together with the recommendations of the National Research Council (NRC)[3] with respect to statistical interpretation of DNA evidence, have been instrumental in the worldwide acceptance of DNA evidence in the criminal justice system. However, the databases on these 13 loci are largely restricted to broadly defined population groups, such as US White, US Black, Hispanics. Although regional data from the United States have been compiled,[4,5] and allele frequency data from Europe are being made available through web site presentations (http://www.uni-Duesseldorf.de/WWW/MedFak/Serology; http://www.cstl.nist.gov/biotech/strbase) in addition to occasional reports from some worldwide populations,[6] data from ethnically defined populations, particularly isolated populations with smaller effective sizes, are relatively scarce. Therefore, our knowledge remains limited as to whether population-related evolutionary forces, such as population bottlenecks and genetic drift among others, impact on the dynamics of these loci and consequently affects their forensic use in specific populations. This is particularly relevant because, with the NRC[3] recommendation for computing statistical significance of a DNA match becoming standard, the need for empirical estimates of worldwide values of genetic differentiation ($F_{ST}$ or $\theta$)[7] among ethnically defined populations has become urgent.[8–10]

With these rationales, we have studied genetic variation at 9 STR markers, which are a subset of the 13 core forensic loci named above, in over 900 individuals drawn from 20 ethnically defined populations representing five major human groups. The objectives are to: (1) generate a worldwide database of allele and genotype frequencies; (2) test independence of alleles within and across loci in the examined populations; (3) estimate the coefficient of co-ancestry at the global as well as major group levels of population differentiation; (4) examine the genetic relationships amongst the sampled populations; and (5) evaluate average match probabilities with and without adjustments for population substructure effects in the studied populations. Additionally, we also present the statistical power of using these loci for parentage testing by evaluating exclusion probabilities for each population database. Our data indicate that, in general, alleles across the nine studied loci are mutually independent in all populations, and pooling of population data by major geographic groupings introduces a co-efficient of co-ancestry no larger than 3.5%, even for small isolated populations (eg, Native Americans). Consequently, even with population substructure adjustment,[3] the estimated match probabilities do not increase by more than 10-fold compared to the ones predicted under the assumption of strict allelic independence. Evaluation of exclusion probabilities indicate that even in the small isolated populations, use of these nine loci offers an exclusionary power above 99.3%. However, in paternity testing with mother's genotype unknown, and with paternity exclusion confirmed by at least two loci, the power of exclusion could fall below 80% in some isolated populations. Therefore, while this worldwide database validates the use of these nine STR loci for DNA-based forensic identification and parentage testing purposes, supplementation with additional loci is recommended for parentage testing in small populations, particularly when the mother is not available for genotyping.

## Materials and methods

### Population samples

The 20 populations surveyed in this research include: ***Africans***, viz. Sudanese (SUD), Nigerian (NIG), Benin (BEN), South Carolina Blacks (SCB); ***Caucasians***, viz. German (GER), Spanish (SPN), United Arab Emirates (UAE), Brazilian White (BRA); ***Asians***, viz. Chinese (CHN), Japanese (JAP), Kachari (KAC), Thai (THA), Kampuchean (KAM); ***Native Americans***, viz. Dogrib (DOG), Ngöbé (NGB), Wounan (WON), Bri Bri (BRI), Pehuenche (PEH); and ***Oceanic***, viz., Samoan (SAM), Papua New Guinea Highlanders (PNG). These populations are globally distributed and representative of large (African, Caucasian and Asian) and small, isolated (Native American and Oceanic) populations known to have undergone recent population bottlenecks. Nigerian, Benin, Sudanese, South Carolina Black, German, Spanish, Arabs from UAE, Brazilian Whites, Chinese, Japanese, Thai, Kampuchean, derive their names from their countries or regions of origins, and are representatives of the broadly defined ancestral groups to which they belong. The Kachari are a Tibetoburman speaking Mongoloid group from Northeast India. Of the American Indian groups, the NaDene speaking Dogrib population is distributed in the Northwest territories of Canada; the Bri Bri from Costa Rica and the Ngöbé from Panama are Chibcha speaking groups; the Wounan from Panama are Chocoan speakers; and the Pehuenche are a group of Araucanian Indians from Chile. Among the Oceanic groups, Samoans are a Polynesian population distributed over the independent nation of Samoa and the US territory of American Samoa; the New Guinea Highlanders are sampled from the Central Highlands of Papua New Guinea. Further details of these populations are also found elsewhere.[11–16]

### DNA analysis

We have used the Profiler Plus kit from Applied Biosystems, which is designed for co-amplification of the nine STR loci. Multiplex PCR amplification of these loci, viz., D3S1358, HumvWA, HumFGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820 and the amelogenin locus was conducted following the protocol in the AmpF1STR Profiler Plus PCR manual,[2] with the only modification of using a 25 $\mu$l PCR reaction volume instead of the 50 $\mu$l as described in the manual. The amplified products were separated on an ABI 377 DNA sequencer. GeneScan 3.1, AmpF1STR Profiler Plus Template and Genotyper 2.5 (Applied Biosystems) software were used for sizing and genotyping.

### Statistical analysis

As the studied loci are autosomal co-dominant, allele frequencies were computed by gene counting.[17] Three tests were used for testing conformity with Hardy–Weinberg proportion of genotype frequencies, viz., exact test for multiallelic loci,[18] log likelihood method,[19] and the homozygosity test.[20] The levels of significance for each test statistic were evaluated through 10 000 replicates of permutations of the observed alleles within each database. As the results of these three tests were in general congruent, we have reported in the text only the levels of significance of the exact test, since this is the most powerful of the three test procedures.[21]

Mutual independence of alleles was tested by two test statistics, each of which utilized the nine-locus genotypes of individuals from each population. The first test statistic is the variance of the number of heterozygous loci across the individual DNA profiles in each database. This test statistic ($s_k^2$) detects the presence of linkage disequilibria across loci,[22,23] which in the context of these unlinked loci, signifies the presence of population substructure within each database. The observed value of $s_k^2$ was compared with its 95% confidence limit estimate based on the assumption of mutual independence of alleles, analytically computed by the methods as described in Brown *et al*[22] and Chakraborty.[23] The second test is based on the distribution of the number of shared alleles between all possible pairs of nine-locus DNA profiles of individuals within each population database. The expected distribution of allele sharing was analytically evaluated based on the theory described in Chakraborty and Jin.[24] Concordance of the observed and expected distributions of allele sharing is the indicator of mutual independence of alleles, relevant for forensic application of such databases.

Estimates of co-ancestry measures were obtained by apportionment analysis of gene diversity and allele size variance, conducted first at the level of geographic grouping of populations (five groups, as mentioned earlier), and second, by using two level substructuring (among five groups, and between-populations within each group), using the theory of Chakraborty *et al*,[25] which is an extension of the AMOVA analysis,[26] adapted for microsatellite loci. The levels of significance of $G_{ST}$ estimates from these computations were determined by the permutation test (10 000 replications).

Average match probability and exclusion probabilities for parentage testing were computed by using the computational formulae as listed in Chakraborty *et al.*[10] For match probability evaluation, impact of possible population substructure effects within each database (judged to be non-significant for each individual population) was examined by computing a weighted conditional match probability (as shown in Appendix 1, since this computational formula is not explicitly available in the literature).

## Results and discussion

The allele frequencies at the nine studied loci are presented in Appendix 2. Occasionally, some DNA samples could not be optimally amplified at some loci, and consequently, sample sizes differ to some extent from one locus to the other. One possible reason for non-amplification could be sequence-variation at the primer-binding site. However, frequencies of null alleles resulting from such phenomena are rare and do not affect the validity of these loci in forensic analysis.[27] Nonetheless, the allele frequency distributions show that each STR locus is substantially polymorphic in the worldwide populations. This is also reflected in summary measures of genetic variation, viz., number of alleles, allele size variance and heterozygosity, which are presented in Table 1. This indicates that the levels of variation at the nine STR loci are high in all populations, irrespective of their effective sizes. Even though the larger continental populations (eg, the populations of African, Caucasian, and Asian descent) show a somewhat larger level of variation, the reduction of genetic diversity in the smaller isolated groups (eg, the Native Americans and the Oceanic populations) appears to be marginally small. It may be argued that this reduced genetic variation could be an artifact of the small sample size of these populations, eg, only 99 Oceanic individuals were sampled compared to the 291 Asians. It is known that average heterozygosity and allele size variance are not strikingly affected by sample size differences of this order.[28,29] However, the number of segregating alleles is sensitive to such sample size effects. To account for this, we computed the expected average number of alleles (as described in [30]) that would have been observed if 99 individuals were sampled from each of the five major geographic groups of populations. From this analysis, we obtain the average number of alleles ranging from 7.8 (among the Native Americans and Oceanic populations) to 10.2 (among the Africans), with the Caucasians and Asians having intermediate allele numbers, 9.4 and 9.3, respectively. Thus, the somewhat reduced levels of diversity at these nine STR loci, among the Native American and Oceanic populations, are not due to their sample size differences, but are rather reflections of genetic drift operating more actively in these populations.

Tests for conformity of genotype frequencies with HWE, performed by the exact test,[18] showed only nine significant departures from equilibrium out of a total of 180 locus-population combinations. Of these, seven (Brazilian at D13S1379, Sudanese at FGA, Chinese at D8S1179, D13S317, D7S820, Ngöbé at D21S11 and Pehunche at D8S1179) were at 5% level of significance, and two (New

**Table 1** Summary statistics of within population variation at nine STR loci in 20 global populations

| | | *Average (SE) over nine loci* | |
|---|---|---|---|
| *Population (No. of individuals)* | *Number of alleles* | *Allele size variance* | *Expected heterozygosity* |
| ***African*** | | | |
| Sudanese (46) | 9.1 (1.2) | 3.46 (0.77) | 0.813 (0.02) |
| Nigerian (46) | 9.4 (1.2) | 2.87 (0.61) | 0.794 (0.02) |
| Benin (51) | 9.2 (1.1) | 2.90 (0.62) | 0.792 (0.02) |
| S.C. Black (48) | 8.9 (1.1) | 3.12 (0.69) | 0.797 (0.03) |
| Pooled (191) | 12.0 (1.6) | 3.11 (0.66) | 0.800 (0.02) |
| ***Caucasian*** | | | |
| German (49) | 8.4 (0.7) | 2.63 (0.44) | 0.814 (0.02) |
| Spanish (46) | 8.6 (0.8) | 2.72 (0.40) | 0.807 (0.02) |
| United Arab Emirates (53) | 8.6 (0.9) | 2.83 (0.55) | 0.811 (0.02) |
| Brazilian (81) | 9.4 (0.8) | 2.94 (0.50) | 0.817 (0.02) |
| Pooled (229) | 10.8 (1.2) | 2.81 (0.47) | 0.814 (0.02) |
| ***Asian*** | | | |
| Chinese (103) | 8.7 (0.9) | 2.61 (0.42) | 0.802 (0.02) |
| Japanese (47) | 8.6 (0.7) | 3.08 (0.66) | 0.799 (0.02) |
| Kachari (54) | 8.9 (0.8) | 3.05 (0.56) | 0.816 (0.02) |
| Thai (48) | 8.7 (1.2) | 2.86 (0.43) | 0.810 (0.02) |
| Kampuchean (39) | 7.9 (0.8) | 2.56 (0.31) | 0.806 (0.01) |
| Pooled (291) | 11.0 (1.2) | 2.82 (0.45) | 0.809 (0.01) |
| ***Native American*** | | | |
| Dogrib (48) | 5.9 (0.6) | 2.39 (0.38) | 0.744 (0.03) |
| Ngöbé (22) | 5.6 (0.7) | 2.15 (0.71) | 0.633 (0.06) |
| Wounan (22) | 6.4 (0.7) | 2.68 (0.64) | 0.724 (0.04) |
| Bri Bri (43) | 6.6 (0.7) | 2.64 (0.73) | 0.733 (0.04) |
| Pehuenche (37) | 6.8 (0.7) | 3.06 (0.87) | 0.732 (0.04) |
| Pooled (172) | 8.3 (0.9) | 2.74 (0.64) | 0.753 (0.03) |
| ***Oceanic*** | | | |
| Samoan (48) | 7.7 (0.5) | 2.41 (0.32) | 0.785 (0.01) |
| New Guinea Highlander (51) | 6.4 (0.5) | 2.53 (0.54) | 0.755 (0.01) |
| Pooled (99) | 8.3 (0.9) | 2.72 (0.42) | 0.791 (0.01) |

Note: DOS version of software used in the computation of this study is available from the authors upon request.

Guinea Highlander at D13S1379 and Sudanese at D21S11) were at 1% level of significance. Overall, the proportion of discordances (9 out of 180) exactly conforms to the nominal level of significance (5%), indicating a general agreement with HW proportions of genotype frequencies in the entire dataset.

In order to examine whether HWE expectations hold at the level of geographic populations, we performed a similar analysis of exact test[18] on the pooled samples within each of the five major groups. Of the 45 locus-group combinations, eight significant deviations were observed. The Africans and the Caucasians were at HWE at all loci; the Asians and the Oceanians showed departure at a single locus each, D5S818 locus ($P$=0.026) and D13S317 ($P$=0.007), respectively. However, among the Native Americans, six of the nine loci were significantly different from the HWE expectations (vWA, $P$=0.049; D8S1179, $P$=0.013; D21S11, $P$=0.005; D18S51, $P$=0.020; D5S818, $P$=0.004; and D7S820, $P$=0.015).

These results, together, suggest that for these nine STR loci, the assumption of HWE holds reasonably well for anthropologically defined populations. Further, when ethnic groups are pooled as geographic and/or broadly defined entities, in general, the larger continental and cosmopolitan populations still adhere to the expectations of HWE. However, groupings of isolated populations even of common ancestral origin, such as the Native Americans, exhibit the presence of population substructure, which could be attributed to the effects of genetic drift resulting from relative isolation and smaller population sizes.

Summary statistics of two tests of mutual independence of the nine STR loci are shown in Table 2. When each multi-locus genotype occurs only once in a sample, the summed number of heterozygous loci is a sufficient statistic for testing the hypothesis of mutual independence of loci.[8] Therefore, the test statistic, $s_k^2$ (the variance of the number of heterozygous loci across individuals, in their nine-locus genotype profiles), used for testing the hypothesis of mutual independence of all loci contains all information in a database of multi-locus genotypes. For all 20 populations, the observed values of $s_k^2$ are within their respective 95% confidence limits, supporting agreement with the hypothesis of mutual independence of loci. This conclusion is also corroborated by the test of conformity of observed and expected number of alleles shared between all pairs of individuals within each population (last two columns of Table 2). Thus, there is no evidence of non-random association of alleles across loci in any of the 20 populations examined in this study. Absence of nonrandom association of alleles at these loci is also evident at the level of major groupings of populations (see Table 2). In addition, allele-sharing data reveals another important population genetic characteristic, which is not readily observed in tabulations of allele frequencies. The last two columns of Table 2 show that individuals, who are members of smaller isolated populations, share more alleles in their multilocus genotype profiles than do the individuals from larger populations. This larger sharing of alleles is, nonetheless, in expectation of random combination of alleles in their genotypes (as seen from the conformity of observed and expected). Thus, the larger allele sharing in Native Americans and Oceanic populations is consistent with their reduced genetic variation (Table 1).

Tables 3 and 4 provide summary results of gene diversity analyses of the nine STR loci. For geographic populations within each of the five major groups, we have evaluated the coefficient of gene diversity $G_{ST}$, which is effectively equivalent to the coefficient of coancestry, $\theta$, based on gene diversity and allele size variance separately.[31] Although in the context of evolutionary relationships of populations, allele size variance based estimates are preferred, gene diversity based estimates are more relevant for forensic applications. Nevertheless, data presented in Table 3 establishes two important points. First, for all major groups of populations, estimates of $\theta < 3\%$ are adequate, as suggested in the NRC report. Second, the levels of significance, obtained by a permutation-based method,[10] indicate that even small values of $\theta$ can be statistically significant. In other words, even when two databases from two different samples from the same population show statistically significant differences of allele frequencies, such observations do

**Table 2** Tests of multi-locus independence of allele frequencies in 20 global populations

| | | Mean (SD) number of shared alleles | |
|---|---|---|---|
| *Population* | *$S_k^2$ (95% CI)* | *Observed* | *Expected* |
| Sudanese | 1.22 (0.89–2.11) | 5.0 (1.6) | 5.4 (1.7) |
| Nigerian | 1.37 (0.77–1.85) | 5.6 (1.7) | 5.7 (1.7) |
| Benin | 1.90 (0.88–1.97) | 5.5 (1.7) | 5.8 (1.8) |
| S.C. Black | 1.05 (0.85–1.97) | 5.4 (1.7) | 5.5 (1.7) |
| ***African Pooled*** | 1.39 (1.15–1.72) | 5.5 (1.8) | 5.5 (1.7) |
| German | 1.33 (0.77–1.81) | 5.3 (1.8) | 5.4 (1.7) |
| Spanish | 1.09 (0.77–1.85) | 5.4 (1.8) | 5.6 (1.7) |
| United Arab Emirates | 1.45 (0.82–1.84) | 5.4 (1.7) | 5.5 (1.7) |
| Brazilian | 1.24 (0.95–1.81) | 5.1 (1.7) | 5.3 (1.7) |
| ***Caucasian Pooled*** | 1.26 (1.10–1.60) | 5.3 (1.8) | 5.3 (1.7) |
| Chinese | 1.62 (1.09–1.90) | 5.4 (1.7) | 5.6 (1.8) |
| Japanese | 1.41 (0.82–1.93) | 5.6 (1.8) | 5.7 (1.8) |
| Kachari | 1.20 (0.82–1.83) | 5.2 (1.8) | 5.4 (1.7) |
| Thai | 1.67 (0.80–1.88) | 5.3 (1.8) | 5.5 (1.7) |
| Kampuchean | 1.68 (0.80–2.04) | 5.5 (1.7) | 5.7 (1.8) |
| ***Asian Pooled*** | 1.52 (1.20–1.66) | 5.4 (1.7) | 5.4 (1.7) |
| Dogrib | 2.22 (1.00–2.30) | 6.6 (1.8) | 6.7 (1.8) |
| Ngöbé | 1.78 (0.75–2.73) | 7.0 (1.4) | 7.5 (1.6) |
| Wounan | 1.69 (0.70–2.54) | 6.4 (1.6) | 6.7 (1.7) |
| Bri Bri | 1.62 (0.97–2.31) | 6.6 (1.8) | 6.6 (1.7) |
| Pehuenche | 2.20 (0.87–2.26) | 6.7 (1.9) | 6.7 (1.7) |
| ***Native American Pooled*** | 2.03 (1.37–2.08) | 6.3 (1.9) | 6.2 (1.7) |
| Samoan | 0.85 (0.82–1.92) | 6.1 (1.7) | 6.1 (1.8) |
| New Guinea Highlander | 1.25 (1.02–2.28) | 6.4 (1.8) | 6.6 (1.8) |
| ***Oceanic Pooled*** | 1.13 (1.12–1.97) | 5.9 (1.9) | 5.9 (1.8) |

Note: $S_k^2$ = Variance of the number of heterozygous loci in nine-locus genotype, computed over all individuals in the population. The number of shared alleles was evaluated by pairwise comparisons of nine-locus genotypes for all possible pairs of individuals within the population.

**Table 3** Estimates of coefficient of gene differentiation ($G_{ST}$) among populations for five major groups of humans based on nine STR loci

| | Based on gene diversity | | Based on allele size variance | |
|---|---|---|---|---|
| *Population groups* | *$G_{ST}$ (H) in %* | *Prob.* | *$G_{ST}$ (V) in %* | *Prob.* |
| African | 0.18±0.16 | 0.045 | 0.80±0.44 | 0.006 |
| Caucasian | 0.22±0.07 | 0.011 | 0.21±0.31 | 0.148 |
| Asian | 0.48±0.10 | $<10^{-4}$ | 0.45±0.33 | 0.028 |
| Native American | 4.07±0.53 | $<10^{-4}$ | 4.97±1.10 | $<10^{-4}$ |
| Oceanic | 2.70±0.63 | $<10^{-4}$ | 9.20±3.80 | $<10^{-4}$ |

**Table 4** Gene diversity analysis of 20 global populations sub-divided as five major groups and sub-populations within each group

| | *Between groups* | | *Between populations within group* | | | |
|---|---|---|---|---|---|---|
| *Locus* | $G_{gt}$ *(H) in %* | $G_{gt}$ *(V) in %* | $G_{sg}$ *(H) in %* | *Prob.* | $G_{sg}$ *(V) in %* | *Prob.* |
| D3S1358 | 2.66 | 7.22 | 1.65 | $<10^{-4}$ | 1.07 | 0.0045 |
| VWA | 4.29 | 0.58 | 1.79 | $<10^{-4}$ | 5.03 | $<10^{-4}$ |
| FGA | 1.94 | 5.07 | 1.28 | $<10^{-4}$ | 6.60 | $<10^{-4}$ |
| D8S1179 | 1.87 | 3.81 | 1.22 | $<10^{-4}$ | 0.59 | 0.0457 |
| D21S11 | 2.46 | 3.97 | 1.62 | $<10^{-4}$ | 1.69 | 0.0001 |
| D18S51 | 1.62 | 3.04 | 1.38 | $<10^{-4}$ | 2.02 | $<10^{-4}$ |
| D5S818 | 3.68 | 10.64 | 1.28 | $<10^{-4}$ | 1.68 | 0.0002 |
| D13S317 | 6.19 | 10.81 | 2.04 | $<10^{-4}$ | 2.14 | $<10^{-4}$ |
| D7S820 | 2.22 | 6.24 | 0.71 | 0.0002 | 2.09 | $<10^{-4}$ |
| Average | 2.97 | 5.33 | 1.44 | $<10^{-4}$ | 2.86 | $<10^{-4}$ |
| SE | 0.50 | 1.06 | 0.13 | | 0.83 | |

not compromise forensic calculations, because such departures can be taken into account by invoking values of θ as suggested in the NRC report.

Table 4 illustrates another aspect of the gene diversity analysis. The estimates of $G_{ST}$ for between populations within a major group are smaller than among the major groups of geographic populations. This provides empirical support for the notion that establishing STR databases based on broad definitions of populations is adequate for use in forensic analysis.[8,32]

Figure 1 shows a neighbour-joining tree[33] of the genetic affinities amongst the 20 populations based on the chord distance,[34] which has been demonstrated to generate reliable tree topologies.[35] We have also estimated the phylogenetic relationships based on Nei's standard genetic distance,[36] which showed very similar topologies and bootstrap values compared with the chord distance (data not shown). A notable feature of the network tree is that, in general, populations within a major geographic or racial group have clustered together. For example, all of the populations of African ancestry are proximally placed, as are the populations of Caucasian/European and Asian origins, respectively. Interestingly, all of the five Native American groups are located on the same branch. An exception is the position of the Samoans, whose branch lies between the Africans and the Caucasians. Based on the known ethno-history and affinity of this population,[37,38] one would expect the Samoans to cluster with other Asian populations. However, the bootstrap values supporting the Samoan branch are rather low and thus this anomalous observation is most likely due to the limited number of markers used. It should also be noted that in a previous study on South-east Asian and Oceanic populations, using a separate set of nine STRs and five Y-specific STR loci in the principal component analysis, the Samoans were an outlier compared to the majority of the South-east Asians.[16]

In Table 5, we illustrate the power of the battery of the nine loci for forensic and parentage testing applications. In general, the nine loci have adequate discriminatory power for forensic identification of individuals, as well as sufficient exclusionary power for parentage analysis. As expected, with adjustment of population substructure effect (ie, with non-zero values of θ), the match probabilities are reduced to some extent. However, for all populations the average match probability is well below their respective current population sizes, reflecting global rarity of nine-locus DNA profiles based on these nine loci. In other words, a somewhat reduced level of genetic variation in isolated populations (such as the Native Americans and the Oceanic populations) does not compromise the use of these STR loci for the purpose of human identification. As shown in the last four columns of Table 5, these loci are also adequate for parentage testing. With the criterion of exclusion based on at least one locus, and with data on the mother–child pair, the exclusion probability exceeds 99.3% in all popula-

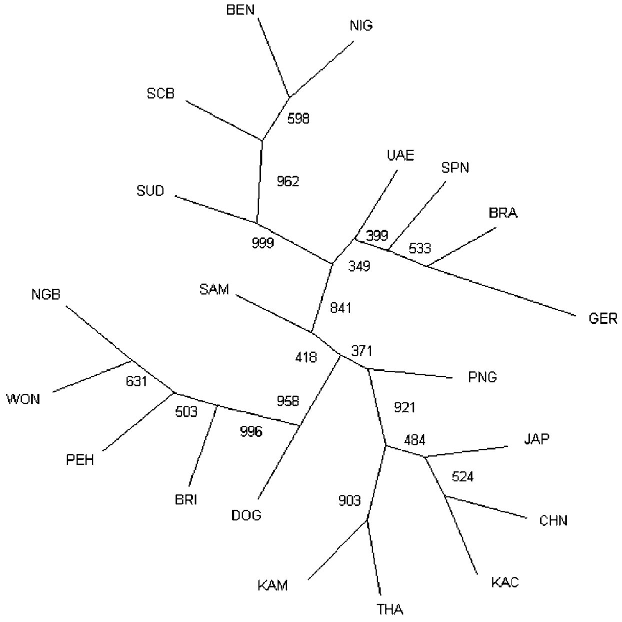

**Figure 1** A neighbour-joining tree based on chord distances. The abbreviated population names are the same as those mentioned in the Materials and methods section. Bootstrap values indicate the degree of support of 1000 replicates for each branch point.

**Table 5** Match probability and paternity exclusion probability with the combined testing of nine STR loci in global populations

| | | | | *Exclusion probability in %* | | | |
|---|---|---|---|---|---|---|---|
| | | | | *With M, C* data and exclusion based on* | | *With data on C only and exclusion based on* | |
| | | *Match probability with* | | | | | |
| *Population* | *θ=0* | *θ=0.01* | *θ=0.03* | *⩾1 locus* | *⩾2 loci* | *⩾1 locus* | *⩾2 loci* |
| Sudanese | $8.3\times10^{-12}$ | $1.8\times10^{-11}$ | $7.0\times10^{-11}$ | 99.990 | 99.813 | 99.654 | 96.787 |
| Nigerian | $3.7\times10^{-11}$ | $7.3\times10^{-11}$ | $2.5\times10^{-10}$ | 99.979 | 99.659 | 99.417 | 95.179 |
| Benin | $5.2\times10^{-11}$ | $9.9\times10^{-11}$ | $3.3\times10^{-10}$ | 99.974 | 99.610 | 99.335 | 94.700 |
| S.C. Black | $2.0\times10^{-11}$ | $4.2\times10^{-11}$ | $1.6\times10^{-10}$ | 99.984 | 99.733 | 99.540 | 95.948 |
| ***African Pooled*** | $1.3\times10^{-11}$ | $2.7\times10^{-11}$ | $1.0\times10^{-10}$ | 99.987 | 99.777 | 99.605 | 96.422 |
| German | $1.2\times10^{-11}$ | $2.3\times10^{-11}$ | $8.6\times10^{-11}$ | 99.987 | 99.787 | 99.600 | 96.465 |
| Spanish | $2.0\times10^{-11}$ | $3.9\times10^{-11}$ | $1.4\times10^{-10}$ | 99.984 | 99.736 | 99.517 | 95.896 |
| United Arab Emirates | $1.4\times10^{-11}$ | $2.8\times10^{-11}$ | $1.0\times10^{-10}$ | 99.986 | 99.766 | 99.567 | 96.224 |
| Brazilian | $6.2\times10^{-12}$ | $1.3\times10^{-11}$ | $5.1\times10^{-11}$ | 99.991 | 99.836 | 99.686 | 97.065 |
| ***Caucasian Pooled*** | $6.2\times10^{-12}$ | $1.3\times10^{-11}$ | $5.1\times10^{-11}$ | 99.991 | 99.835 | 99.685 | 97.056 |
| Chinese | $2.9\times10^{-11}$ | $5.5\times10^{-11}$ | $1.8\times10^{-10}$ | 99.981 | 99.704 | 99.458 | 95.543 |
| Japanese | $4.3\times10^{-11}$ | $8.1\times10^{-11}$ | $2.6\times10^{-10}$ | 99.976 | 99.643 | 99.359 | 94.928 |
| Kachari | $9.9\times10^{-12}$ | $2.0\times10^{-11}$ | $7.6\times10^{-11}$ | 99.988 | 99.799 | 99.623 | 96.614 |
| Thai | $1.3\times10^{-11}$ | $2.7\times10^{-11}$ | $9.9\times10^{-11}$ | 99.987 | 99.777 | 99.591 | 96.366 |
| Kampuchean | $3.2\times10^{-11}$ | $6.3\times10^{-11}$ | $2.1\times10^{-10}$ | 99.979 | 99.673 | 99.403 | 95.203 |
| ***Asian Pooled*** | $1.1\times10^{-11}$ | $2.2\times10^{-11}$ | $7.9\times10^{-11}$ | 99.988 | 99.792 | 99.607 | 96.521 |
| Dogrib | $1.3\times10^{-9}$ | $2.1\times10^{-9}$ | $5.3\times10^{-9}$ | 99.884 | 98.647 | 98.041 | 87.993 |
| Ngöbé | $1.2\times10^{-7}$ | $1.7\times10^{-7}$ | $3.6\times10^{-7}$ | 99.303 | 94.316 | 93.256 | 70.885 |
| Wounan | $2.9\times10^{-9}$ | $4.8\times10^{-9}$ | $1.2\times10^{-8}$ | 99.844 | 98.289 | 97.483 | 85.577 |
| Bri Bri | $1.5\times10^{-9}$ | $2.4\times10^{-9}$ | $6.3\times10^{-9}$ | 99.885 | 98.611 | 98.065 | 87.788 |
| Pehuenche | $1.5\times10^{-9}$ | $2.5\times10^{-9}$ | $6.6\times10^{-9}$ | 99.887 | 98.642 | 98.135 | 88.243 |
| ***Native American Pooled*** | $3.0\times10^{-10}$ | $5.5\times10^{-10}$ | $1.6\times10^{-9}$ | 99.943 | 99.236 | 98.812 | 91.664 |
| Samoan | $1.6\times10^{-10}$ | $2.8\times10^{-10}$ | $7.9\times10^{-10}$ | 99.955 | 99.398 | 98.962 | 92.713 |
| New Guinea Highlander | $1.2\times10^{-9}$ | $1.8\times10^{-9}$ | $4.5\times10^{-9}$ | 99.883 | 98.685 | 97.963 | 87.920 |
| ***Oceanic Pooled*** | $7.9\times10^{-11}$ | $1.4\times10^{-10}$ | $4.3\times10^{-10}$ | 99.967 | 99.538 | 99.169 | 93.813 |

* M = Mother; C = Child.

tions. Exclusion based on at least two loci offers exclusion probability in excess of 94.3%. However, in motherless cases, for some populations (particularly for the small isolated ones), the exclusion probability falls below 80%. Thus, there may be a need to supplement these loci with additional markers for cases that involve unknown mothers, or more complicated forensic scenarios, eg, DNA mixtures involving two or more samples. In view of our observation that pooling of data from Native Americans produced a considerable degree of departure from HWE (but not generating $F_{ST}/G_{ST}$ above 4.1%, see Table 3), a further degree of conservativeness in forensic use of our data presented in Appendix 2 may be achieved by imposing a minimum threshold allele frequency, a concept advocated in forensic literature.[3,39]

In summary, this report establishes a nine-locus STR database in a globally diverse set of anthropologically defined populations. Analyses of genotype and allele frequency data demonstrate that the assumptions of HWE and multi-locus independence of alleles are globally applicable for the STR loci, and sampling designs generally employed in human genetic surveys provide adequate representations of random samples for DNA typing. Gene diversity analysis reflects that when STR databases are pooled over populations by geographic groupings, population substructure effects can be accounted for with values of θ consistent with the ones recommended in NRC report (ie, $\theta < 1\%$ for all cosmopolitan populations, and $\approx 3\%$ for small isolated populations). Finally, with regard to the power of discrimination and exclusion probability, data presented here also show that a reduced level of genetic variation in smaller and isolated populations does not substantially compromise the utility of these loci.

## Acknowledgements

*This work was funded by US Public Health Service Research grants GM 41399, GM 45861 and GM 58545 from the US National Institutes of Health, and grant NIJ-98-LB-VX-K019 from the US National Institute of Justice.*

## References

1 Budowle B, Moretti TR, Niezgoda SJ, Brown BL: CODIS and PCR-based short tandem repeat loci: Law enforcement tools; in: *Second European Symposium on Human Identification 1998*. Madison, WI: Promega Corporation, 1998; pp 73–88.

2 *AmpFLSTR Profiler Plus PCR amplification kit user's manual*. Foster City, CA: PE Applied Biosystems, 1998.

3 *The evaluation of forensic evidence*. National Research Council Report II. Washington DC: National Academy Press, 1996.

4 Budowle B, Shea B, Niezgoda S, Chakraborty R: CODIS STR loci data from 41 sample populations. *J Forensic Sci* 2000; **46**: 453–489.

5 Budowle B, Chakraborty R: Population variation at the CODIS core short tandem repeat loci in Europeans. *Legal Med* 2001; **3**: 29–33.

6 Dutta R, Reddy BM, Chattopadhyay P, Kashyap VK, Sun G, Deka R: Patterns of genetic diversity at the nine forensically approved STR loci in the Indian populations. *Hum Biol* 2002; **74**: 33–49.

7 Wright S: The genetical structure of populations. *Ann Eugen* 1951; **15**: 323–354.

8 Evett IW, Weir BS: *Interpreting DNA Evidence*. Sunderland, MA: Sinauer, 1998.

9 Balding DJ: When can a DNA profile be regarded as unique? *Science and Justice* 1999; **39**: 257–260.

10 Chakraborty R, Stivers DN, Su B, Zhong Y, Budowle B: The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing loci. *Electrophoresis* 1999; **20**: 1682–1696.

11 Barrantes R, Smouse PE, Mohrenweiser HW *et al*: 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* 1990; **46**: 63–84.

12 Deka R, Jin L, Shriver MD *et al*: Population genetics of dinucleotide (dG-dT)n polymorphism in world populations. *Am J Hum Genet* 1995; **56**: 461–474.

13 Deka R, Shriver MD, Yu LM *et al*: Genetic variation at twenty-three microsatellite loci in sixteen human populations. *J Genet* 1999; **78**: 99–121.

14 Kamboh MI, Williams ER, Law J *et al*: Molecular basis of a unique African variant (A-IV 5) of human apolipoprotein A-IV and its significance in lipid metabolism. *Genet Epidemiol* 1992; **9**: 379–388.

15 Kolman CJ, Bermingham E: Mitochondrial and nuclear DNA diversity in Choc and Chibcha Amerinds of Panama. *Genetics* 1997; **147**: 1289–1302.

16 Parra E, Saha N, Soemantri AG *et al*: Genetic Variation at 9 microsatellite loci in Asian and Pacific populations. *Hum Biol* 1999; **71**: 757–779.

17 Li CC: *First Course in Population Genetics*. Pacific Grove, CA: Boxwood, 1976.

18 Guo SW, Thompson EA: Performing the exact test of Hardy-Weinberg proportion of multiple alleles. *Genetics* 1992; **48**: 361–372.

19 Weir BS: *Genetic Data Analysis–II*. Sunderland, MA: Sinauer, 1996.

20 Hartl DL, Clark AG: *Principles of Population Genetics*. Sunderland, MA: Sinauer, 1997.

21 Chakraborty R, Zhong Y: Statistical power of an exact test of Hardy Weinberg proportions of genotypic data at a multiallelic locus. *Hum Hered* 1994; **44**: 1–9.

22 Brown AHD, Feldman MW, Nevo E: Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics* 1980; **96**: 523–536.

23 Chakraborty R: Detection of nonrandom association of alleles from the distribution of the number of heterozygous loci in a sample. *Genetics* 1984; **108**: 719–731.

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

25 Chakraborty R, Jin L, Deka R, Kimmel M: Extent and pattern of gene diversity at microsatellite loci: Implications for disease-gene association studies. *Am J Hum Genet* 2001; **69**: A262.

26 Excoffier L, Smouse PE, Quattro JM: Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial DNA restriction data. *Genetics* 1992; **131**: 479–491.

27 Budowle B, Masibay A, Anderson SJ *et al*: STR primer concordance study. *Forensic Sci International* 2001; **124**: 47–54.

28 Nei M: Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 1978; **89**: 583–590.

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

30 Chakraborty R, Smouse PE, Neel JV: Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. *Am J Hum Genet* 1988; **43**: 709–725.

31 Chakraborty R, Danker-Hopfe H: Analysis of population structure: A comparative study of different estimators of Wright's fixation indices; in: Rao CR, Chakraborty R (eds): *Handbook of Statistics–Vol. 8*. Amsterdam, North-Holland: Elsevier Science Publishers B.V. 1991; pp 203–254.

32 Chakraborty R, Kidd KK: The utility of DNA typing in forensic work. *Science* 1991; **254**: 1735–1739.

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

34 Cavalli-Sforza LL, Edwards AWF: Phylogenetic analysis: models and estimation procedures. *Am J Hum Genet* 1967; **19**: 233–257.

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

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

37 Bellwood P: *Man's Conquest of the Pacific: The Prehistory of South-east Asia and Oceania*. New York: Oxford University Press, 1978.

38 Su B, Underhill P, Martinson J *et al*: Polynesian origins: insights from the Y chromosome. *Proc Natl Acad Sci USA* 2000; **97**: 8225–8228.

39 Budowle B, Monson KL, Chakraborty R: Estimating minimum allele frequencies for DNA profile frequency estimates for PCR-based loci. *Int J Legal Med* 1996; **108**: 173–176.

40 Li CC, Chakravarti A: 1994 DNA profile similarity in a subdivided population. *Hum Hered* 1994; **44**: 100–109.

## **Appendix 1** Average match probability for an autosomal codominant multi-allelic locus in a substructured population

DNA profiles of two individuals are declared to be a match if they exhibit identical genotypes. Thus, for an autosomal codominant multi-allelic locus, with $k$ segregating alleles ($A_1$, $A_2$, …, $A_k$), the average match probability can be written as

$$P_m = \sum_i \Pr(A_iA_i, A_iA_i) + \sum_{i<j}\sum_j \Pr(A_iA_j, A_iA_j) \qquad \text{(A1)}$$

where Pr ($A_iA_i$, $A_iA_i$) and Pr ($A_iA_j$, $A_iA_j$) represent the probabilities of both of the two individuals being homozygote ($A_iA_i$) and heterozygote ($A_iA_j$) for the same set of alleles ($i$, $j$=1, 2, …, $k$). In a substructured population, these individual terms are not simply the squares of an individual's (respective) genotype frequencies, since under a mutation-drift equilibrium model[1], the joint probability of observing $t_i$ copies of the $i$-th allele ($A_i$) for any subset of {$i$=1, 2, …, $k$} alleles in a sample of $t_.$=Σ $t_i$ alleles is given by [8]

$$\Pr(\prod_i A_i^{t_i}) = \frac{\Gamma(\gamma_\bullet)}{\Gamma(\gamma_\bullet + t_\bullet)} \prod_i \frac{\Gamma(\gamma_i + t_i)}{\Gamma(\gamma_i)} \qquad \text{(A2)}$$

where $\gamma_i = p_i(1 - \theta)/\theta$, $\gamma_. = \Sigma\ \gamma_i = (1-\theta)/\theta$, $p_i$=frequency of the $i$-th allele ($A_i$), averaged over all sub-populations, θ=the measure of co-ancestry of individuals (equivalent to Wright's $F_{ST}$[7]), and Γ(•) is the Gamma function.

It is easy to show that this general Dirichlet distribution yields the expected frequencies of individual genotype frequencies as

$$\Pr(A_iA_i) = p_i[(1 \quad \theta)p_i + \theta] \qquad \text{(A3)}$$

and

$$\Pr(A_iA_j) = 2p_ip_j(1 \quad \theta) \tag{A4}$$

However, the application of the same general formula (equation A2) leads to the closed form expressions of the individual terms of the average match probability as

$$\begin{aligned}\Pr(A_iA_i, A_iA_i) &= \Pr(A_i^{\,4})\\ &= \frac{\Gamma(\gamma_\bullet)}{\Gamma(\gamma_\bullet + 4)} \times \frac{\Gamma(\gamma_i + 4)}{\Gamma(\gamma_i)}\\ &= \frac{\gamma_i(\gamma_i + 1)(\gamma_i + 2)(\gamma_i + 3)}{\gamma_\bullet(\gamma_\bullet + 1)(\gamma_\bullet + 2)(\gamma_\bullet + 3)}\\ &= \frac{p_i[\theta + (1 \quad \theta)p_i][2\theta + (1 \quad \theta)p_i][3\theta + (1 \quad \theta)p_i]}{(1 + \theta)(1 + 2\theta)}\end{aligned} \tag{A5}$$

and,

$$\begin{aligned}\Pr(A_iA_j, A_iA_j) &= 4\Pr(A_i^{\,2}A_j^{\,2})\\ &= 4 \times \frac{\Gamma(\gamma_\bullet)}{\Gamma(\gamma_\bullet + 4)} \times \frac{\Gamma(\gamma_i + 2)}{\Gamma(\gamma_i)} \times \frac{\Gamma(\gamma_j + 2)}{\Gamma(\gamma_j)}\\ &= \frac{4\gamma_i(\gamma_i + 1)\gamma_j(\gamma_j + 2)}{\gamma_\bullet(\gamma_\bullet + 1)(\gamma_\bullet + 2)(\gamma_\bullet + 3)}\\ &= \frac{4(1 \quad \theta)p_ip_j[\theta + (1 \quad \theta)p_i][\theta + (1 \quad \theta)p_j]}{(1 + \theta)(1 + 2\theta)}\end{aligned} \tag{A6}$$

Substituting these terms in equation (A1) and with some algebraic simplifications, we obtain a closed form expression for the average match probability given by

$$P_m = \frac{2\theta^2(1 + \theta) + \theta(1 \quad \theta)(4 + 5\theta)m_2 + 2\theta(1 \quad \theta)^2 m_3 + (1 \quad \theta)^3(2m_2^2 \quad m_4)}{(1 + \theta)(1 + 2\theta)} \tag{A7}$$

where $m_r$ is the $r$-th moment of the allele frequency distribution at the locus, represented by $m_r = \sum_{pi}{}^{r}$, for $r$=2, 3, and 4. Note that this expression (equation A7) is different from the one obtained by Li and Chakravarti,[40] since they incorrectly replaced the individual terms of the right hand side of equation (A1) by the squares of expressions (A3) and (A4), respectively. In contrast, as shown above, a non-zero correlation of alleles between individuals, induced by the population substructure effect, violates this approximation (see equations A5 and A6), leading to the error of Li and Chakravarti's final equation. However, when θ=0 (ie, the population is not substructured), our final equation (A7) agrees with the corresponding expression of Li and Chakravarti,[40] namely $P_m$=2$m_2^2$ – $m_4$.

**Appendix 2** Allele frequencies at nine STR loci in 20 global populations

| *Locus* | *Repeat* | | | | | | | | | | *Populations** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Name* | *number* | *SUD* | *NIG* | *BEN* | *SCB* | *GER* | *SPN* | *UAE* | *BRA* | *CHN* | *JAP* | *KAC* | *THA* | *KAM* | *DOG* | *NGB* | *WON* | *BRI* | *PEH* | *SAM* | *PNG* |
| D3S1358 | | | | | | | | | | | | | | | | | | | | | |
| | 11 | 0.0 | 1.0 | 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 | 0.0 | 1.0 | 0.0 |
| | 12 | 0.0 | 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 | 0.0 | 0.0 | 0.0 |
| | 13 | 3.1 | 0.0 | 8.8 | 1.0 | 0.0 | 1.1 | 0.9 | 0.0 | 0.0 | 0.0 | 0.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 14 | 11.2 | 13.5 | 27.5 | 13.3 | 13.3 | 5.3 | 4.7 | 9.9 | 5.3 | 1.0 | 6.5 | 3.1 | 1.3 | 20.4 | 0.0 | 2.2 | 22.1 | 2.6 | 2.1 | 0.0 |
| | 15 | 22.4 | 31.3 | 30.4 | 37.8 | 23.5 | 28.7 | 19.8 | 27.8 | 33.5 | 33.3 | 30.6 | 17.7 | 26.3 | 48.0 | 87.0 | 52.2 | 39.5 | 47.4 | 26.0 | 31.7 |
| | 16 | 24.5 | 31.3 | 25.5 | 27.6 | 26.5 | 25.5 | 33.0 | 22.2 | 35.0 | 37.3 | 34.3 | 27.1 | 35.0 | 17.3 | 13.0 | 34.8 | 23.3 | 46.1 | 39.6 | 28.8 |
| | 17 | 24.5 | 17.7 | 5.9 | 15.3 | 21.4 | 12.8 | 25.5 | 22.2 | 19.9 | 20.6 | 20.4 | 38.5 | 28.8 | 13.3 | 0.0 | 10.9 | 15.1 | 3.9 | 18.8 | 17.3 |
| | 18 | 13.3 | 5.2 | 0.0 | 5.1 | 15.3 | 23.4 | 16.0 | 14.8 | 5.3 | 5.9 | 6.5 | 13.5 | 8.8 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 10.4 | 22.1 |
| | 19 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.2 | 0.0 | 1.2 | 1.0 | 2.0 | 0.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.1 | 0.0 |
| No. of Chr. | | 98 | 96 | 102 | 98 | 98 | 94 | 106 | 162 | 206 | 102 | 108 | 96 | 80 | 98 | 46 | 46 | 86 | 76 | 96 | 104 |
| vWA | | | | | | | | | | | | | | | | | | | | | |
| | 13 | 0.0 | 0.0 | 1.0 | 3.1 | 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 | 0.0 |
| | 14 | 7.3 | 8.7 | 5.9 | 7.3 | 10.2 | 16.0 | 4.7 | 10.5 | 21.8 | 14.0 | 23.1 | 27.1 | 24.4 | 18.4 | 0.0 | 0.0 | 1.2 | 0.0 | 25.0 | 0.0 |
| | 15 | 13.5 | 30.4 | 24.5 | 19.8 | 18.4 | 9.6 | 15.1 | 10.5 | 1.0 | 3.0 | 4.6 | 3.1 | 2.6 | 0.0 | 2.2 | 2.2 | 7.0 | 3.9 | 17.7 | 0.0 |
| | 16 | 31.3 | 23.9 | 26.5 | 24.0 | 19.4 | 28.7 | 26.4 | 23.5 | 23.3 | 17.0 | 17.6 | 11.5 | 19.2 | 33.7 | 56.5 | 67.4 | 66.3 | 60.5 | 13.5 | 19.2 |
| | 17 | 19.8 | 17.4 | 21.6 | 19.8 | 23.5 | 26.6 | 27.4 | 29.6 | 23.8 | 27.0 | 32.4 | 25.0 | 15.4 | 48.0 | 37.0 | 15.2 | 17.4 | 23.7 | 25.0 | 35.6 |
| | 18 | 17.7 | 8.7 | 10.8 | 14.6 | 16.3 | 12.8 | 19.8 | 20.4 | 20.4 | 30.0 | 11.1 | 21.9 | 19.2 | 0.0 | 4.3 | 15.2 | 8.1 | 5.3 | 13.5 | 29.8 |
| | 19 | 9.4 | 6.5 | 6.9 | 6.3 | 10.2 | 6.4 | 6.6 | 3.7 | 8.3 | 6.0 | 9.3 | 8.3 | 19.2 | 0.0 | 0.0 | 0.0 | 0.0 | 6.6 | 5.2 | 12.5 |
| | 20 | 1.0 | 1.1 | 2.9 | 5.2 | 2.0 | 0.0 | 0.0 | 1.2 | 1.5 | 3.0 | 1.9 | 3.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.9 |
| | 21 | 0.0 | 2.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 22 | 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 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| No. of Chr. | | 96 | 92 | 102 | 96 | 98 | 94 | 106 | 162 | 206 | 100 | 108 | 96 | 78 | 98 | 46 | 46 | 86 | 76 | 96 | 104 |
| FGA | | | | | | | | | | | | | | | | | | | | | |
| | 16 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 17 | 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 | 2.3 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 18 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 | 1.1 | 2.8 | 0.6 | 1.0 | 5.2 | 2.8 | 1.0 | 2.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 18.2 | 0.0 | 0.0 | 1.0 | 5.2 | 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 | 0.0 |
| | 19 | 4.3 | 7.6 | 8.8 | 5.2 | 13.3 | 6.5 | 6.6 | 3.7 | 5.8 | 5.2 | 12.0 | 5.2 | 5.1 | 3.1 | 0.0 | 9.1 | 1.2 | 7.9 | 2.1 | 33.7 |

Continued

**Appendix 2** *(Continued)*

| *Locus* | *Repeat* | *Populations** | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Name* | *number* | *SUD* | *NIG* | *BEN* | *SCB* | *GER* | *SPN* | *UAE* | *BRA* | *CHN* | *JAP* | *KAC* | *THA* | *KAM* | *DOG* | *NGB* | *WON* | *BRI* | *PEH* | *SAM* | *PNG* |
| | 20 | 6.4 | 3.3 | 2.9 | 5.2 | 19.4 | 8.7 | 9.4 | 17.9 | 4.4 | 6.3 | 8.3 | 4.2 | 3.8 | 11.5 | 0.0 | 0.0 | 9.3 | 7.9 | 1.0 | 9.6 |
| | 20.2 | 0.0 | 0.0 | 2.0 | 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 | 0.0 | 0.0 |
| | 21 | 11.7 | 10.9 | 15.7 | 12.5 | 19.4 | 14.1 | 10.4 | 14.2 | 14.6 | 11.5 | 11.1 | 13.5 | 15.4 | 12.5 | 4.5 | 0.0 | 0.0 | 9.2 | 0.0 | 0.0 |
| | 21.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.8 | 3.1 | 2.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 22 | 18.1 | 18.5 | 16.7 | 16.7 | 10.2 | 21.7 | 16.0 | 16.0 | 18.9 | 17.7 | 16.7 | 20.8 | 21.8 | 16.7 | 4.5 | 2.3 | 20.9 | 1.3 | 3.1 | 4.8 |
| | 22.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.2 | 0.0 | 1.2 | 0.5 | 0.0 | 0.9 | 3.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 23 | 17.0 | 18.5 | 15.7 | 17.7 | 16.3 | 13.0 | 18.9 | 16.7 | 14.6 | 20.8 | 17.6 | 18.8 | 19.2 | 9.4 | 4.5 | 11.4 | 20.9 | 13.2 | 28.1 | 21.2 |
| | 23.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.9 | 0.0 | 2.8 | 0.0 | 1.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 24 | 11.7 | 17.4 | 16.7 | 18.8 | 10.2 | 16.3 | 19.8 | 16.7 | 20.9 | 22.9 | 10.2 | 7.3 | 15.4 | 17.7 | 25.0 | 27.3 | 22.1 | 10.5 | 30.2 | 12.5 |
| | 24.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.1 | 0.0 | 0.0 | 2.4 | 0.0 | 0.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.3 | 0.0 | 0.0 | 0.0 |
| | 25 | 8.5 | 12.0 | 6.9 | 9.4 | 7.1 | 10.9 | 10.4 | 7.4 | 13.1 | 6.3 | 10.2 | 7.3 | 7.7 | 27.1 | 36.4 | 29.5 | 15.1 | 25.0 | 21.9 | 15.4 |
| | 25.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 4.2 | 2.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 26 | 5.3 | 3.3 | 7.8 | 3.1 | 2.0 | 3.3 | 1.9 | 3.7 | 1.0 | 3.1 | 3.7 | 6.3 | 2.6 | 2.1 | 11.4 | 9.1 | 7.0 | 21.1 | 12.5 | 2.9 |
| | 26.2 | 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 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 27 | 5.3 | 2.2 | 3.9 | 2.1 | 1.0 | 1.1 | 1.9 | 0.6 | 1.0 | 1.0 | 0.9 | 2.1 | 0.0 | 0.0 | 13.6 | 9.1 | 1.2 | 3.9 | 1.0 | 0.0 |
| | 27.2 | 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 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 28 | 6.4 | 5.4 | 1.0 | 1.0 | 0.0 | 0.0 | 0.9 | 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 |
| | 29 | 4.3 | 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 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 31 | 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 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 32.2 | 0.0 | 0.0 | 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 | 0.0 | 0.0 |
| No. of Chr. | | 94 | 92 | 102 | 96 | 98 | 92 | 106 | 162 | 206 | 96 | 108 | 96 | 78 | 96 | 44 | 44 | 86 | 76 | 96 | 104 |
| D8S1179 | | | | | | | | | | | | | | | | | | | | | |
| | 7 | 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 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 8 | 0.0 | 0.0 | 0.0 | 0.0 | 2.0 | 3.3 | 0.0 | 2.5 | 0.0 | 0.0 | 0.9 | 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 | 1.0 | 2.2 | 0.9 | 1.2 | 0.0 | 1.0 | 0.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 10 | 3.1 | 0.0 | 2.0 | 3.0 | 4.1 | 6.5 | 6.6 | 6.8 | 12.6 | 14.0 | 8.3 | 9.4 | 14.1 | 4.0 | 2.2 | 6.5 | 0.0 | 0.0 | 8.3 | 0.0 |
| | 11 | 1.0 | 1.1 | 0.0 | 3.0 | 9.2 | 8.7 | 8.5 | 6.8 | 11.2 | 6.0 | 8.3 | 14.6 | 12.8 | 7.0 | 13.0 | 13.0 | 31.4 | 23.1 | 4.2 | 10.4 |
| | 12 | 14.6 | 11.7 | 15.7 | 11.0 | 19.4 | 12.0 | 9.4 | 14.2 | 9.2 | 13.0 | 5.6 | 10.4 | 15.4 | 23.0 | 17.4 | 15.2 | 11.6 | 10.3 | 1.0 | 15.1 |
| | 13 | 24.0 | 25.5 | 14.7 | 18.0 | 31.6 | 19.6 | 23.6 | 31.5 | 23.3 | 22.0 | 20.4 | 11.5 | 24.4 | 21.0 | 32.6 | 54.3 | 33.7 | 43.6 | 33.3 | 36.8 |
| | 14 | 29.2 | 37.2 | 33.3 | 39.0 | 20.4 | 30.4 | 25.5 | 24.1 | 15.5 | 20.0 | 23.1 | 17.7 | 7.7 | 25.0 | 4.3 | 6.5 | 10.5 | 17.9 | 31.3 | 14.2 |
| | 15 | 18.8 | 21.3 | 24.5 | 22.0 | 8.2 | 14.1 | 22.6 | 10.5 | 16.5 | 17.0 | 23.1 | 18.8 | 23.1 | 20.0 | 30.4 | 4.3 | 8.1 | 2.6 | 15.6 | 21.7 |
| | 16 | 9.4 | 3.2 | 6.9 | 4.0 | 3.1 | 3.3 | 2.8 | 2.5 | 9.7 | 6.0 | 8.3 | 13.5 | 1.3 | 0.0 | 0.0 | 0.0 | 4.7 | 0.0 | 4.2 | 0.9 |
| | 17 | 0.0 | 0.0 | 2.0 | 0.0 | 1.0 | 0.0 | 0.0 | 0.0 | 1.5 | 0.0 | 0.9 | 4.2 | 1.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.1 | 0.9 |
| | 18 | 0.0 | 0.0 | 1.0 | 0.0 | 0.0 | 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 |
| | 19 | 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 | 0.0 | 0.0 | 2.6 | 0.0 | 0.0 |
| No. of Chr. | | 96 | 94 | 102 | 100 | 98 | 92 | 106 | 162 | 206 | 100 | 108 | 96 | 78 | 100 | 46 | 46 | 86 | 78 | 96 | 106 |
| D21S11 | | | | | | | | | | | | | | | | | | | | | |
| | 23.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| | 24.2 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.1 | 0.0 | 0.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 25 | 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 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 25.2 | 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 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 26 | 0.0 | 1.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 9.3 | 0.0 | 0.0 | 0.0 |
| | 27 | 8.2 | 4.2 | 2.9 | 5.1 | 3.1 | 0.0 | 0.0 | 1.9 | 0.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 | 0.0 |
| | 28 | 9.2 | 28.1 | 34.3 | 18.4 | 13.3 | 10.6 | 18.9 | 19.8 | 5.3 | 4.1 | 8.3 | 6.3 | 5.0 | 3.1 | 0.0 | 2.2 | 1.2 | 2.7 | 26.0 | 1.0 |
| | 28.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.9 | 3.1 | 0.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 29 | 25.5 | 16.7 | 16.7 | 19.4 | 17.3 | 25.5 | 24.5 | 17.9 | 24.8 | 32.7 | 25.0 | 18.8 | 18.8 | 6.1 | 8.3 | 6.5 | 24.4 | 9.5 | 32.3 | 36.5 |
| | 29.2 | 0.0 | 1.0 | 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 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 30 | 21.4 | 15.6 | 10.8 | 14.3 | 23.5 | 27.7 | 15.1 | 20.4 | 29.1 | 28.6 | 15.7 | 21.9 | 28.8 | 21.4 | 54.2 | 45.7 | 47.7 | 33.8 | 16.7 | 40.4 |
| | 30.2 | 0.0 | 2.1 | 0.0 | 5.1 | 8.2 | 2.1 | 2.8 | 3.7 | 1.5 | 0.0 | 4.6 | 3.1 | 6.3 | 7.1 | 0.0 | 0.0 | 0.0 | 1.4 | 0.0 | 0.0 |
| | 31 | 4.1 | 10.4 | 12.7 | 13.3 | 8.2 | 5.3 | 10.4 | 6.2 | 11.7 | 7.1 | 3.7 | 6.3 | 11.3 | 12.2 | 14.6 | 4.3 | 1.2 | 6.8 | 6.3 | 9.6 |
| | 31.2 | 10.2 | 2.1 | 2.9 | 5.1 | 11.2 | 9.6 | 10.4 | 11.7 | 7.3 | 9.2 | 9.3 | 9.4 | 5.0 | 15.3 | 0.0 | 13.0 | 0.0 | 4.1 | 6.3 | 5.8 |
| | 32 | 1.0 | 3.1 | 1.0 | 0.0 | 1.0 | 1.1 | 0.9 | 0.6 | 4.9 | 1.0 | 3.7 | 5.2 | 2.5 | 0.0 | 4.2 | 0.0 | 2.3 | 0.0 | 0.0 | 0.0 |
| | 32.2 | 7.7 | 6.3 | 3.9 | 5.1 | 9.2 | 10.6 | 11.3 | 8.0 | 9.7 | 10.2 | 19.4 | 21.9 | 18.8 | 22.4 | 10.4 | 17.4 | 10.5 | 24.3 | 9.4 | 3.8 |
| | 33 | 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 | 0.0 | 0.0 | 0.0 | 1.9 |
| | 33.2 | 5.1 | 2.1 | 3.9 | 3.1 | 4.1 | 5.3 | 3.8 | 6.8 | 3.4 | 3.1 | 9.3 | 7.3 | 2.5 | 12.2 | 2.1 | 6.5 | 3.5 | 17.6 | 2.1 | 0.0 |
| | 34 | 1.0 | 1.0 | 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 | 0.0 | 0.0 |
| | 34.2 | 0.0 | 0.0 | 0.0 | 1.0 | 1.0 | 1.1 | 0.0 | 1.2 | 0.0 | 0.0 | 0.0 | 0.0 | 1.3 | 0.0 | 6.3 | 4.3 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 35 | 4.1 | 3.1 | 5.9 | 7.1 | 0.0 | 0.0 | 0.0 | 0.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 36 | 1.0 | 2.1 | 1.0 | 1.0 | 0.0 | 0.0 | 0.9 | 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 |
| | 37 | 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 | 0.0 | 0.0 | 0.0 | 0.0 |
| No. of Chr. | | 98 | 96 | 102 | 98 | 98 | 94 | 106 | 162 | 206 | 98 | 108 | 96 | 80 | 98 | 48 | 46 | 86 | 74 | 96 | 104 |

Continued

**Appendix 2** *(Continued)*

| *Locus* | *Repeat* | *Populations** | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Name* | *number* | *SUD* | *NIG* | *BEN* | *SCB* | *GER* | *SPN* | *UAE* | *BRA* | *CHN* | *JAP* | *KAC* | *THA* | *KAM* | *DOG* | *NGB* | *WON* | *BRI* | *PEH* | *SAM* | *PNG* |
| D18S51 | | | | | | | | | | | | | | | | | | | | | |
| | 9 | 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 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 10 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.1 | 0.0 | 3.1 | 0.0 | 0.0 | 0.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 10.2 | 1.1 | 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 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 11 | 1.1 | 1.1 | 1.0 | 2.0 | 1.0 | 1.1 | 6.6 | 1.9 | 0.0 | 1.0 | 0.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 | 0.0 |
| | 11.2 | 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 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 12 | 14.9 | 9.6 | 3.9 | 8.2 | 13.3 | 13.8 | 7.5 | 12.3 | 1.9 | 2.1 | 2.8 | 6.3 | 7.5 | 6.3 | 0.0 | 0.0 | 0.0 | 8.1 | 2.1 | 0.0 |
| | 13 | 13.8 | 2.1 | 3.9 | 6.1 | 16.3 | 16.0 | 16.0 | 10.5 | 22.3 | 25.0 | 22.2 | 8.3 | 11.3 | 12.5 | 15.9 | 10.9 | 16.3 | 5.4 | 2.1 | 7.8 |
| | 13.2 | 1.1 | 1.1 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.2 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 14 | 6.4 | 5.3 | 2.0 | 6.1 | 16.3 | 13.8 | 22.6 | 14.2 | 24.3 | 24.0 | 20.4 | 17.7 | 13.8 | 28.1 | 13.6 | 23.9 | 12.8 | 32.4 | 16.7 | 32.4 |
| | 14.2 | 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 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 15 | 11.7 | 19.1 | 15.7 | 23.5 | 19.4 | 14.9 | 14.2 | 13.6 | 21.4 | 13.5 | 19.4 | 29.2 | 31.3 | 41.7 | 2.3 | 28.3 | 14.0 | 9.5 | 25.0 | 14.7 |
| | 15.2 | 0.0 | 1.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.9 | 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 |
| | 16 | 10.6 | 9.6 | 18.6 | 12.2 | 11.2 | 19.1 | 7.5 | 9.9 | 9.7 | 9.4 | 13.0 | 17.7 | 17.5 | 0.0 | 2.3 | 4.3 | 12.8 | 4.1 | 4.2 | 5.9 |
| | 17 | 19.1 | 22.3 | 23.5 | 9.2 | 8.2 | 8.5 | 10.4 | 13.6 | 8.3 | 5.2 | 3.7 | 6.3 | 8.8 | 3.1 | 36.4 | 10.9 | 14.0 | 13.5 | 24.0 | 22.5 |
| | 18 | 5.3 | 10.6 | 13.7 | 16.3 | 7.1 | 7.4 | 4.7 | 9.9 | 3.4 | 8.3 | 3.7 | 4.2 | 1.3 | 0.0 | 6.8 | 8.7 | 8.1 | 9.5 | 7.3 | 10.8 |
| | 19 | 9.6 | 8.5 | 9.8 | 9.2 | 4.1 | 1.1 | 3.8 | 7.4 | 2.4 | 3.1 | 3.7 | 2.1 | 5.0 | 6.3 | 4.5 | 4.3 | 5.8 | 4.1 | 14.6 | 2.9 |
| | 20 | 1.1 | 6.4 | 2.9 | 3.1 | 0.0 | 2.1 | 2.8 | 3.7 | 1.9 | 3.1 | 3.7 | 3.1 | 2.5 | 2.1 | 9.1 | 0.0 | 3.5 | 2.7 | 2.1 | 2.0 |
| | 21 | 1.1 | 2.1 | 2.0 | 1.0 | 0.0 | 0.0 | 1.9 | 0.0 | 1.0 | 1.0 | 3.7 | 1.0 | 1.3 | 0.0 | 0.0 | 2.2 | 2.3 | 1.4 | 0.0 | 0.0 |
| | 22 | 2.1 | 0.0 | 1.0 | 2.0 | 2.0 | 1.1 | 0.0 | 0.0 | 2.4 | 0.0 | 2.8 | 2.1 | 0.0 | 0.0 | 9.1 | 2.2 | 10.5 | 9.5 | 0.0 | 1.0 |
| | 23 | 0.0 | 1.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.9 | 0.0 | 1.0 | 3.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 24 | 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 | 2.2 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 25 | 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 | 0.0 | 0.0 | 0.0 | 1.0 | 0.0 |
| | 27 | 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 | 0.0 | 0.0 | 0.0 | 0.0 |
| No. of Chr. | | 94 | 94 | 102 | 98 | 98 | 94 | 106 | 162 | 206 | 96 | 108 | 96 | 80 | 96 | 44 | 46 | 86 | 74 | 96 | 102 |
| D5S818 | | | | | | | | | | | | | | | | | | | | | |
| | 6 | 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 | 0.0 | 0.0 | 0.0 | 1.0 |
| | 7 | 0.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.2 | 1.9 | 0.0 | 0.9 | 1.0 | 2.5 | 29.6 | 10.4 | 15.2 | 8.1 | 19.7 | 0.0 | 0.0 |
| | 8 | 9.4 | 4.2 | 4.9 | 8.0 | 0.0 | 0.0 | 0.0 | 1.9 | 0.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.1 | 0.0 | 0.0 | 0.0 | 0.0 | 1.9 |
| | 9 | 3.1 | 2.1 | 0.0 | 0.0 | 2.0 | 2.1 | 5.7 | 1.9 | 4.4 | 10.0 | 6.5 | 6.3 | 2.5 | 0.0 | 20.8 | 17.4 | 20.9 | 11.8 | 0.0 | 0.0 |
| | 10 | 11.5 | 15.6 | 8.8 | 9.0 | 12.2 | 4.3 | 14.2 | 7.4 | 19.9 | 17.0 | 18.5 | 24.0 | 23.8 | 7.1 | 0.0 | 0.0 | 0.0 | 1.3 | 17.7 | 24.0 |
| | 11 | 24.0 | 21.9 | 19.6 | 21.0 | 35.7 | 45.7 | 32.1 | 33.3 | 34.0 | 30.0 | 34.3 | 25.0 | 31.3 | 52.0 | 58.3 | 37.0 | 29.1 | 43.4 | 11.5 | 26.9 |
| | 12 | 28.1 | 33.3 | 34.3 | 40.0 | 37.8 | 28.7 | 34.0 | 37.7 | 27.0 | 17.0 | 22.2 | 26.0 | 12.5 | 2.0 | 8.3 | 23.9 | 40.7 | 14.5 | 30.2 | 31.7 |
| | 13 | 21.9 | 19.8 | 28.4 | 21.0 | 11.2 | 18.1 | 13.2 | 14.8 | 12.6 | 22.0 | 15.7 | 17.7 | 27.5 | 9.2 | 0.0 | 6.5 | 1.2 | 9.2 | 32.3 | 14.4 |
| | 14 | 2.1 | 2.1 | 3.9 | 1.0 | 1.0 | 1.1 | 0.9 | 1.2 | 0.0 | 0.0 | 1.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.1 | 0.0 |
| | 15 | 0.0 | 0.0 | 0.0 | 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 | 4.2 | 0.0 |
| | 16 | 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 | 0.0 | 0.0 | 0.0 | 1.0 | 0.0 |
| No. of Chr. | | 96 | 96 | 102 | 100 | 98 | 94 | 106 | 162 | 206 | 100 | 108 | 96 | 80 | 98 | 48 | 46 | 86 | 76 | 96 | 104 |
| D13S317 | | | | | | | | | | | | | | | | | | | | | |
| | 6 | 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 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 7 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 | 0.0 | 0.9 | 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 |
| | 8 | 13.5 | 2.1 | 1.0 | 5.0 | 14.3 | 20.2 | 16.0 | 9.9 | 23.3 | 30.6 | 22.2 | 36.5 | 33.3 | 0.0 | 0.0 | 0.0 | 1.2 | 0.0 | 5.2 | 44.2 |
| | 9 | 3.1 | 0.0 | 2.0 | 2.0 | 8.2 | 2.1 | 5.7 | 8.6 | 13.6 | 11.2 | 13.9 | 10.4 | 12.8 | 32.7 | 65.2 | 39.1 | 31.4 | 25.7 | 21.9 | 1.0 |
| | 10 | 0.0 | 2.1 | 1.0 | 2.0 | 3.1 | 6.4 | 5.7 | 8.6 | 18.4 | 6.1 | 13.0 | 9.4 | 11.5 | 12.2 | 6.5 | 10.9 | 15.1 | 18.9 | 5.2 | 1.9 |
| | 11 | 20.8 | 24.5 | 30.4 | 25.0 | 34.7 | 21.3 | 18.9 | 32.1 | 23.8 | 30.6 | 31.5 | 25.0 | 23.1 | 23.5 | 6.5 | 13.0 | 9.3 | 8.1 | 27.1 | 19.2 |
| | 12 | 45.8 | 51.1 | 43.1 | 52.0 | 28.6 | 30.9 | 35.8 | 22.8 | 14.1 | 17.3 | 14.8 | 13.5 | 15.4 | 22.4 | 8.7 | 13.0 | 8.1 | 17.6 | 35.4 | 27.9 |
| | 13 | 10.4 | 12.8 | 15.7 | 11.0 | 8.2 | 16.0 | 11.3 | 13.0 | 4.9 | 3.1 | 2.8 | 4.2 | 3.8 | 9.2 | 10.9 | 15.2 | 29.1 | 10.8 | 3.1 | 5.8 |
| | 14 | 5.2 | 6.4 | 6.9 | 3.0 | 2.0 | 3.2 | 5.7 | 4.9 | 1.5 | 1.0 | 1.9 | 0.0 | 0.0 | 0.0 | 2.2 | 8.7 | 5.8 | 18.9 | 2.1 | 0.0 |
| | 15 | 1.0 | 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 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| No. of Chr. | | 96 | 94 | 102 | 100 | 98 | 94 | 106 | 162 | 206 | 98 | 108 | 96 | 78 | 98 | 46 | 46 | 86 | 74 | 96 | 104 |
| D7S820 | | | | | | | | | | | | | | | | | | | | | |
| | 6 | 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 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 7 | 0.0 | 1.1 | 2.0 | 1.0 | 1.0 | 3.2 | 2.8 | 1.2 | 0.0 | 0.0 | 0.0 | 1.0 | 2.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 8 | 15.2 | 23.4 | 23.5 | 21.4 | 11.2 | 13.8 | 14.2 | 13.0 | 14.6 | 15.6 | 25.0 | 11.5 | 13.8 | 10.4 | 0.0 | 13.0 | 0.0 | 1.4 | 8.3 | 27.9 |
| | 9 | 15.2 | 10.6 | 5.9 | 12.2 | 19.4 | 11.7 | 4.7 | 8.6 | 4.4 | 3.1 | 6.5 | 6.3 | 8.8 | 0.0 | 0.0 | 2.2 | 1.2 | 2.7 | 9.4 | 8.7 |
| | 10 | 43.5 | 34.0 | 39.2 | 31.6 | 28.6 | 34.0 | 35.8 | 30.2 | 14.6 | 17.7 | 24.1 | 13.5 | 22.5 | 22.9 | 34.1 | 15.2 | 27.9 | 25.7 | 25.0 | 17.3 |
| | 11 | 18.5 | 21.3 | 20.6 | 20.4 | 22.4 | 20.2 | 22.6 | 21.6 | 37.9 | 36.5 | 26.9 | 41.7 | 35.0 | 31.3 | 27.3 | 39.1 | 24.4 | 48.6 | 20.8 | 22.1 |
| | 12 | 6.5 | 6.4 | 7.8 | 13.3 | 15.3 | 16.0 | 17.0 | 18.5 | 23.8 | 19.8 | 13.9 | 22.9 | 13.8 | 30.2 | 36.4 | 21.7 | 45.3 | 20.3 | 17.7 | 22.1 |
| | 13 | 1.1 | 2.1 | 0.0 | 0.0 | 1.0 | 1.1 | 2.8 | 4.9 | 4.9 | 6.3 | 2.8 | 3.1 | 2.5 | 5.2 | 2.3 | 8.7 | 1.2 | 1.4 | 8.3 | 1.9 |
| | 14 | 0.0 | 0.0 | 1.0 | 0.0 | 1.0 | 0.0 | 0.0 | 1.9 | 0.0 | 1.0 | 0.9 | 0.0 | 1.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 10.4 | 0.0 |
| No. of Chr. | | 92 | 94 | 102 | 98 | 98 | 94 | 106 | 162 | 206 | 96 | 108 | 96 | 80 | 96 | 44 | 46 | 86 | 74 | 96 | 104 |

* SUD (Sudanese), NIG (Nigerian), BEN (Benin), SCB (South Carolina Black), GER (German), SPN (Spanish), UAE (United Arab Emirates), BRA (Brazilian White), CHN (Chinese), JAP (Japanese), KAC (Kachari), THA (Thailand), KAM (Kampuchean), DOG (Dogrib), NGB (Ngöbé), WOU (Wounam), BRI (Bri Bri), PEH (Pehuenche), SAM (Samoan), PNG (Papua New Guinea Highlanders).