5-1-2015

Genetic Diversity in Sioux Indians of South Dakota with 21 Autosomal STR Loci and Their Forensic Utility

Nicole C. Grubb
University of North Texas Health Science Center at Fort Worth, nicolecg12@yahoo.com

Grubb, Nicole. Genetic Diversity in Sioux Indians of South Dakota with 21 Autosomal STR Loci and Their Forensic Utility. Master of Science (Biomedical Sciences, Forensic Genetics). May 2015. 23 pp., 28 tables, 1 figure, 11 titles.

Short Tandem Repeats are the current standard markers in DNA Forensics. They have been extensively studied in cosmopolitan populations, but not so in small isolated populations, which tend to have unique societal structures impacting their genetic diversity. This project examined 21 autosomal STR loci in Sioux Indians and used NIST data of the same loci on four major U.S. populations to make the comparisons. The study concluded that the Sioux Indians have a reduced genetic diversity at these loci, but this reduction did not compromise their utility for forensic applications for human identification. Genetic diversity analyses of individual Sioux Indian tribes suggested that a pooled database is appropriate for DNA forensic applications.

GENETIC DIVERSITY IN SIOUX INDIANS OF SOUTH DAKOTA WITH 21 AUTOSOMAL STR LOCI AND THEIR FORENSIC UTILITY

Nicole Grubb, B.S.

APPROVED:

_____

Major Professor

_____

Committee Member

_____

Committee Member

_____

University Member

_____

Chair, Department of Molecular and Medical Genetics

_____

Dean, Graduate School of Biomedical Sciences

GENETIC DIVERSITY IN SIOUX INDIANS OF SOUTH DAKOTA WITH 21 AUTOSOMAL STR LOCI AND THEIR FORENSIC UTILITY

THESIS

Presented to the Graduate Council of the

Graduate School of Biomedical Sciences

University of North Texas

Health Science Center at Fort Worth

in Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

By

Nicole Grubb, B.S.

Fort Worth, TX

May 2015

ACKNOWLEDGEMENTS

I wish to express my thanks to my major professor, Dr. Ranajit Chakraborty, for all of the assistance and guidance throughout my time at the University of North Texas Health Science Center. I would also like to thank all of the professors who have provided leadership and support to me throughout my time as a student. Special thanks to my committee members: Dr. Bruce Budowle, Dr. Robert Barber, and Dr. Laszlo Prokai for their support and guidance. I am extremely grateful for Dr. Laura Gaydosh-Combs, Ms. Jie Sun, and Ms. Shahida Flores for their never ending support and guidance. I would like to thank my fellow colleagues in the Forensic Genetics program for their endless friendship and support over the last two years. Lastly, I wish to express thanks to my family, my fiancé, and loved ones for always supporting me through this time.

TABLE OF CONTENTS

Page

LIST OF TABLES……..……iv

LIST OF FIGURES……..…….....v

CHAPTERS

I. INTRODUCTION……..…….……1

II. METHODOLOGY……..……...…….…..….....4

III. RESULTS AND DISCUSSION……...6

IV. CONCLUSIONS…..…..…….…….....11

APPENDIX……12

REFERENCES……..…….....23

LIST OF TABLES

Table 1: Empirical P-values of HWE Tests, Theta Values, and Chi Square Results ……... 7

Table 2: Heterozygosity Values and Number of Observed Alleles……..... 8

Table 3: Distance Matrix…… 9

Table 4: Forensically Relevant Calculations for Sioux Indians……..……...10

Table 5: D3S1348……..13

Table 6: vWA……...….13

Table 7: D16S539……..13

Table 8: TPOX…….....13

Table 9: CSF1PO……..14

Table 10: D8S1179……...14

Table 11: D21S11…….14

Table 12: D18S51…….15

Table 13: D2S441…….15

Table 14: TH01…….....15

Table 15: D19S433……...16

Table 16: FGA……..16

Table 17: D13S317……...16

Table 18: D22S1045…….17

Table 19: D5S818…….17

Table 20: D7S820……..17

Table 21: SE33……...18

Table 22: D10S1248…….....18

Table 23: D1S1656……19

Table 24: D12S391……19

Table 25: D2S1338……20

Table 26: Random Match Probability Calculation Comparisons……..20

Table 27: Probability of Exclusion with Mother Child Data Calculation Comparisons……...21

Table 28: Probability of Exclusion with Child Data Calculation Comparisons……22

LIST OF FIGURES

Figure 1: Genetic Distance Tree…….8

CHAPTER 1

INTRODUCTION

Short Tandem Repeats (STRs) are currently the gold standard markers of human identification in DNA forensics. Many population studies have been carried out to determine the individualizing power of these markers (1,2). These studies have shown that STRs are diverse at both the individual as well as population levels. In relation to populations of continental USA (such as Caucasians, African-Americans, Hispanics, or US Asians), the cosmopolitan populations have been well studied, whereas some of the smaller, more isolated populations (Native Americans, for example) need further examination. Such an information gap is even more serious for the STR markers that have been introduced more recently.

Such isolated populations tend to have unique societal structures. For example, there are many different tribes of Native Americans, and each tribe is broken down even further into tribal communities. It is of interest to see how this structure impacts the genetic diversity of these populations as compared to the larger populations. There are "…sufficient 'among population' differences [that] exist among Native American tribes to warrant special attention to ensure a representative Native American DNA database for the US" (3). Currently, the CODIS database contains data only on tribes that are in the arctic and subarctic regions of North America, such as Navajo and Apache. One study that provided this information was conducted on three Native Alaskan populations. This study showed that these populations are highly polymorphic, but also showed reduced genetic diversity as compared to the major cosmopolitan populations (4).

Similar results are also available for various other Native American populations in the continental USA and specific isolated populations of the world (1,5).

However, these studies only account for a small portion of the contemporary tribes present in North America. One Native American population that has not had its autosomal STRs examined is the Sioux Indians. HLA haplotypes of the Sioux Indian were examined early on (6). This study showed that there was limited HLA diversity, an element that had been observed among other Native American tribes. It also demonstrated that there were differences of allele prevalence when the Sioux Indian data was compared to other native groups (6). What this means is that one allele could be very frequent in one native group and not seen in another. Y STRs of the Sioux Indian were compared with a Caucasian population, and there were notable differences in allele frequencies (7). This study also illustrated that the Y STR haplotypes of the Sioux Indian formed a distinct cluster when compared to other native groups through phylogenetic analysis (7).

Based on such information, it is of interest to examine whether the Sioux Indian genetic isolation has led to reduced genetic diversity and differences among tribes for other loci as well. Due to marriage within the communities and geographic isolation, it is expected that there would be appreciable genetic difference between tribal communities of the Sioux Indian population for further additional polymorphic loci. However, if the loci chosen for such studies are STR loci, the hypervariability of these loci (due to their inherent large mutation rates), would suggest that even with reduced genetic diversity and geographic isolation, STR DNA profiles in Sioux Indians should provide an adequately efficient power of discrimination for human identification and forensic genetic analysis.

The objective of this research was to address the issues of reduced genetic diversity and geographic isolation, by considering genotype data on 21 autosomal STR loci collected from nine tribal communities of Sioux Indians in South Dakota. The specific aims addressed in this study were: (i) the extent of reduction of genetic diversity in Sioux Indians in comparison with the four major cosmopolitan populations of the U.S. in terms of heterzygosity and observed number of segregating alleles; (ii) the impact of such reduced genetic diversity in Sioux Indians on their population structure (namely, deviation from Hardy Weinberg and Linkage Equilibrium, and genetic distances of the tribal communities); and (iii) the consequences of these features of genetic diversity in Sioux Indians on the utility of this panel of STR markers for forensic applications in the population.

CHAPTER 2

METHODOLOGY

Anonymized genotype data on a sample of 242 presumably unrelated individuals from 9 tribal communities (Cheyenne River, Standing Rock, Flandreau, Lower Brule, Crow Creek, Pine Ridge, Rosebud, Sisseton-Whapeton, and Yankton) of Sioux Indians of South Dakota is available for this research through a collaboration with South Dakota Forensic Laboratory in Pierre, South Dakota. When performing analyses on tribal communities, Flandreau was not studied due to sample size, and some tribal communities were combined together. The Cheyenne River and Standing Rock tribes were combined, as were the Lower Brule and Crow Creek tribes. These communities were grouped together based on the proximity of the reservations to each other (7). The tribal communities will be referred to as CRSR for Cheyenne River and Standing Rock, LBCC for Lower Brule and Crow Creek, PR for Pine Ridge, R for Rosebud, W for Sisseton-Whapeton, and Y for Yankton, with SI representing the pooled population of Sioux Indians. Genotype data from National Institute of Standards and Technology (NIST) on 1,035 individuals (available from www.nist.gov) from the four comparison cosmopolitan populations was also used. This analysis was performed using the GDA software (8) to determine allele frequencies, to conduct tests for Hardy Weinberg Equilibrium (to ascertain that alleles within loci combine at random to form locus-specific genotypes of individuals), and to conduct tests for Linkage Equilibrium (i.e. alleles across all pairs of loci are

independent). The outputs from the GDA software were used to compute the observed and expected heterozygosity. In addition, genetic distances between tribes were estimated using the coancestry measure of identity. An in-house software of empirical tests of contingency table data was used to determine statistical differences between allele frequencies of six tribal communities of Sioux Indians. Computations for random match probability were calculated using the equation listed as A7 in Sun et al. 2003 (5). The paternity exclusion with mother and child data was calculated using the equation from Chakraborty et al. 1996 (9). Calculations for paternity exclusion with child data were performed using the equation $PE=1-4m_2+4m_3-3m_4+2m_2^2$, where m_r is the sum of the r-th power of allele frequencies at a locus (r = 2, 3 and 4).

CHAPTER 3

RESULTS AND DISCUSSION

Results of the Hardy Weinberg Equilibrium (HWE) test (seen in Table 1) were particularly noteworthy. For each individual test, the p-value used was 0.05. However, since 21 tests were performed for HWE tests, a bonferroni adjustment resulted in a p-value of 0.0024. One tribal population showed statistical significance at one locus at this level. However, when data on the tribal communities were pooled, seven loci showed statistical significance for the pooled data of Sioux Indians. These values are bolded in the table. This finding was explored further by examining theta values, both to signify the extent of impact of departure from HWE and to determine if there were significant differences among allele frequencies of the tribal communities. After the chi square analysis, two loci showed significant deviation of allele frequencies across tribes. These values are bolded in the table. The highest theta value observed did correspond to a locus that did not meet Hardy Weinberg Equilibrium. However, overall theta values were not high; in fact, nine values were negative. After studying the Linkage Equilibrium, only 37 out of 210 comparisons passed. These results are based on a p-value of 0.05, after bonferroni adjustment, the p-value was 0.00024. However, when the number of comparisons that included the seven loci that did not meet Hardy Weinberg Equilibrium were counted, 132 out of the 140 of those comparisons failed Linkage Equilibrium.

Locus	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*	*SI*	*Theta*	*P-values from Chi Square Test*
D3S1358	0.0397	0.0777	0.0268	0.0374	0.4017	**0.0018**	**0.0008**	0.0645	0.6427
vWA	0.1940	0.1820	0.0331	0.5710	0.9567	0.2860	0.2306	0.0448	0.4960
D16S539	0.3423	0.0577	0.4973	0.2432	0.2997	0.2345	0.7848	-0.0166	0.9013
CSF1PO	0.4687	0.3670	0.0476	0.5757	0.3809	0.7298	0.0999	-0.0402	0.6148
TPOX	0.7626	0.2299	0.1622	0.6298	0.4811	0.6766	0.0403	-0.0679	0.7833
D8S1179	0.2417	0.7584	0.0105	0.5147	0.1980	0.0487	0.0047	0.0367	0.4165
D21S11	0.1290	0.3308	0.1424	0.3557	0.1051	0.3714	**0.0007**	0.0551	0.2076
D18S51	0.9443	0.7200	0.1289	0.7359	0.7691	0.6089	0.2533	0.0235	0.2214
D2S441	0.1434	0.4856	0.0878	0.0793	0.2629	0.3374	**0.0002**	-0.0232	0.5661
D19S433	0.3367	0.3884	0.7694	0.2149	0.7538	0.4717	0.0345	0.0064	0.1434
TH01	0.3313	0.8202	0.2764	0.1795	0.0748	0.2794	0.0026	0.0314	0.1748
FGA	0.9872	0.6211	0.8486	0.4697	0.1114	0.9349	0.5118	-0.0116	0.2305
D22S1045	0.1163	0.1068	0.6735	0.3632	0.2863	0.9749	0.1029	-0.0041	0.3944
D5S818	0.8225	0.4832	0.2848	0.6822	0.5954	0.3835	**0.0019**	0.0459	0.0448
D13S317	0.9375	0.9436	0.3992	0.2183	0.4627	0.0514	0.1480	-0.0124	**0.0004**
D7S820	0.0702	0.3657	0.2751	0.4910	0.2946	0.2681	0.0052	-0.0115	0.8490
SE33	0.8900	1.0000	0.0084	0.1008	0.5580	0.7746	**0.0002**	0.0264	**$<10^{-4}$**
D10S1248	0.5538	0.1645	0.1373	0.1513	0.5349	0.0176	0.0080	0.0359	0.4654
D1S1656	0.1344	0.9416	0.0612	0.6041	0.0204	0.2686	**0.0005**	0.0356	0.2857
D12S391	0.1716	0.1959	0.0943	0.4754	0.4831	0.0154	**0.0001**	0.1327	0.2830
D2S1338	0.1889	0.6706	0.2214	0.7395	0.9391	0.2375	0.0051	-0.0071	0.3814

Table 1. Empirical p-values of Hardy Weinberg Equlibrium (HWE) tests within the tribal communities and in the pooled data (SI), together with the Theta values, and obtained p-values from Chi Square analysis for allele frequency differences among the tribal communities.

The reduced diversity in the Sioux Indians, compared with the four cosmopolitan populations, was examined by using two summary statistics of genetic variation: heterozygosity and observed number of segregating alleles. These results are displayed in Table 2. Sioux Indians had about four to five percent reduced heterozygosity and about two to three less number of alleles observed on average. Due to the small sample size of the Asian data used, for comparisons of observed number of alleles in Asians versus Sioux Indians, calculations were performed using the method described in Chakraborty et al 1988 (10). This analysis allows for a more direct comparison, and demonstrated that the Sioux Indians had about one less number of

alleles observed on average than the Asian population. For comparisons of heterzygosity values, no such sample size adjustment was needed since heterzygosity estimates are not so sensitive on sample size differences of the magnitude seen in this dataset (n ranging from 97 to 361). In 1999, a study comparing six populations with an Apache Native American population showed similar results (11).

Locus	*Asian (n=97)*	*African American (n=341)*	*Caucasian (n=361)*	*Hispanic (n=236)*	*Sioux Indian (n=242)*
D3S1358	0.7152 (6)	0.7482 (9)	0.7914 (9)	0.7644 (8)	0.62723 (7)
vWA	0.7925 (7)	0.8137 (11)	0.8087 (10)	0.7983 (9)	0.7268 (8)
D16S539	0.7733 (6)	0.7969 (8)	0.7610 (7)	0.7939 (8)	0.7357 (7)
CSF1PO	0.7542 (8)	0.7792 (8)	0.7204 (7)	0.7221 (9)	0.7547 (7)
TPOX	0.6007 (5)	0.7646 (8)	0.6413 (8)	0.6794 (7)	0.6655 (5)
D8S1179	0.8548 (9)	0.8052 (11)	0.8083 (10)	0.8118 (11)	0.7292 (8)
D21S11	0.8165 (12)	0.8492 (20)	0.8299 (16)	0.8393 (16)	0.8309 (11)
D18S51	0.8425 (13)	0.8799 (19)	0.8770 (15)	0.8817 (15)	0.8082 (15)
D2S441	0.7536 (10)	0.7600 (11)	0.7703 (11)	0.7504 (10)	0.6219 (9)
D19S433	0.8011 (11)	0.8563 (14)	0.7705 (15)	0.7960 (13)	0.8443 (12)
TH01	0.6993 (6)	0.7438 (7)	0.7656 (8)	0.7789 (6)	0.5674 (7)
FGA	0.8523 (13)	0.8745 (23)	0.8605 (14)	0.8828 (16)	0.8700 (13)
D22S1045	0.7667 (8)	0.8255 (11)	0.7237 (8)	0.6843 (9)	0.6831 (8)
D5S818	0.8010 (8)	0.7513 (9)	0.6986 (9)	0.7155 (9)	0.7103 (7)
D13S317	0.8059 (6)	0.7059 (7)	0.7848 (8)	0.8344 (8)	0.8203 (7)
D7S820	0.7571 (7)	0.7736 (9)	0.8172 (9)	0.7820 (8)	0.7476 (6)
SE33	0.9440 (22)	0.9320 (41)	0.9493 (38)	0.9420 (36)	0.9125 (25)
D10S1248	0.7773 (7)	0.8049 (11)	0.7596 (9)	0.7547 (9)	0.7244 (8)
D1S1656	0.8455 (12)	0.8664 (15)	0.9004 (15)	0.8869 (15)	0.8656 (14)
D12S391	0.8435 (15)	0.8592 (19)	0.8920 (16)	0.8797 (19)	0.8576 (14)
D2S1338	0.8746 (12)	0.8947 (13)	0.8826 (12)	0.8763 (11)	0.8247 (11)
Average Heterozygosity	0.7939	0.8136	0.8006	0.8026	0.7585
STD	0.0717	0.0591	0.075	0.0721	0.0931
Average Number of Alleles Observed	9.67	13.52	12.09	12.00	9.95
STD	4.04	7.81	6.70	6.53	4.52

Table 2. Expected heterozygosity values and number of observed alleles in the pooled sample of Sioux Indians as compared to those in 4 major cosmopolitan populations of continental USA.

A distance matrix was then using the coancestry measure of identity for the tribes (seen in Table 3). This showed very small values between the tribes. When this information was used

to generate a tree, all the tribes came together at a nodal distance value of 0.0047 (illustrated in Figure 1). This observation indicates that a pooled database is sufficient for the Sioux Indians as long as adjustment for substructuring is performed with a theta value of 0.0047. Note that the National Research Council (NRC), in their recommendation of 1996, suggested use of theta value of 0.03 for small isolated populations (12).

	CRSR	*LBCC*	*PR*	*R*	*W*	*Y*
CRSR	-	0.007764	-0.0005	0.000729	0.002125	-0.00102
LBCC	0.007794	-	0.004098	0.002702	0.008999	0.000462
PR	-0.0005	0.004106	-	-0.00088	0.002601	0.001012
R	0.000729	0.002706	-0.00088	-	0.002504	0.0017
W	0.002127	0.00904	0.002605	0.002508	-	0.004532
Y	-0.00102	0.000462	0.001012	0.001701	0.004543	-

Table 3. Distance Matrix

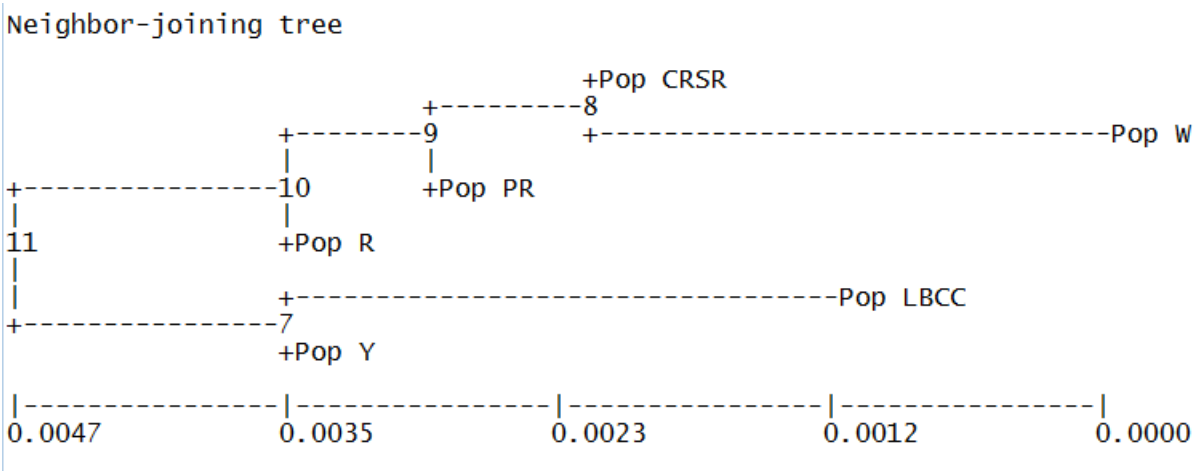

Figure 1. Genetic Distance Tree. CRSR-Cheyenne River and Standing Rock, LBCC-Lower Brule and Crow Creek, PR-Pine Ridge, R-Rosebud, W-Sisseton-Whapeton, and Y-Yankton

Three different calculations were performed using the STR allele frequencies. Tribal allele frequency tables by loci can be found in the appendix. The three calculations performed were average random match probability, paternity exclusion with mother/child data, and

paternity exclusion with child data only. Table 4 shows the results of the calculations for the pooled Sioux Indian data. The random match probability was performed using specific locus theta, an overall theta (0.0173), and no theta (Θ=0). These numbers showed an estimated frequency of 1 in 24.7, 76.2 and 972.7 sextillion individuals. As for the paternity exclusion calculations, 99.99% of individuals can be excluded in a typical paternity test. Similar calculations were performed to compare to the four major populations (Asian, African American, Caucasian, and Hispanic). These comparisons can be seen in a table in the appendix.

Locus	*RMP (Loci Specific Θ)*	*RMP (Θ=0.0173)*	*RMP (Θ=0)*	*PEmc*	*PEc*
D3S1358	0.2330	0.1999	0.1874	0.3885	0.2214
vWA	0.1468	0.1313	0.1217	0.4883	0.3154
D16S539	0.1132	0.1229	0.1132	0.5026	0.3261
CSF1PO	0.1002	0.1094	0.1002	0.5285	0.3499
TPOX	0.1741	0.1838	0.1741	0.4012	0.2428
D8S1179	0.1383	0.1270	0.1171	0.4993	0.3244
D21S11	0.0771	0.0583	0.0503	0.6639	0.4928
D18S51	0.0663	0.0626	0.0525	0.6539	0.4783
D2S441	0.2004	0.2119	0.2004	0.3694	0.2115
D19S433	0.0460	0.0508	0.0431	0.6888	0.5221
TH01	0.2522	0.2404	0.2255	0.3476	0.1792
FGA	0.0314	0.0382	0.0314	0.7345	0.5777
D22S1045	0.1568	0.1668	0.1568	0.4314	0.2676
D5S818	0.1505	0.1314	0.1198	0.4931	0.3113
D13S317	0.0572	0.0653	0.0572	0.6407	0.4656
D7S820	0.1054	0.1147	0.1054	0.5189	0.3416
SE33	0.0230	0.0199	0.0145	0.8208	0.6961
D10S1248	0.1351	0.1232	0.1122	0.5067	0.3253
D1S1656	0.0479	0.0399	0.0329	0.7284	0.5705
D12S391	0.1021	0.0441	0.0369	0.7126	0.5511
D2S1338	0.0532	0.0615	0.0532	0.6553	0.4836
Combined	4.05×10^{-22}	1.31×10^{-22}	1.03×10^{-23}	0.9999	0.9999

Table 4. Forensically relevant calculations (RMP- the random match probability, PEmc-Paternity Exclusion with Mother-child data; PEc- Paternity Exclusion with Child data only) for the Sioux Indians.

CHAPTER 4

CONCLUSIONS

Short tandem repeats (STRs) are very useful in DNA Forensics. However, there is not a lot of population data on STRs for smaller isolated groups, particularly for the expanded panel of STR loci introduced recently (13). Analyses of 21 autosomal STRs in the Sioux Indian population of South Dakota have shown that an overall Sioux Indian database is satisfactory for the purposes of DNA Forensic cases implicating the Sioux Indian population of South Dakota. As expected, this study proved that the Sioux Indian population does have reduced genetic diversity compared with the four major cosmopolitan populations (about four to five percent reduction in heterozygosity and one to two less number of alleles segregating per locus). However, this reduction has not dramatically affected important forensic calculations, such as random match probability and the probability of exclusion. In aggregate, as in the case of the major cosmopolitan populations of continental USA, this expanded panel of 21 STR loci has a comparable level of efficiency for human identification and paternity analyses in Sioux Indians.

APPENDIX

Table 5. D3S1358

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
13	0.013889	0	0.005882	0	0	0
14	0.083333	0.075	0.070588	0.0375	0.041667	0.108696
15	0.458333	0.575	0.541176	0.5875	0.666667	0.413043
16	0.305556	0.225	0.223529	0.2	0.222222	0.369565
17	0.097222	0.1	0.129412	0.1	0.027778	0.086957
18	0.041667	0.025	0.023529	0.0625	0.041667	0.021739
19	0	0	0.005882	0.0125	0	0

Table 6. vWA

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
14	0.013889	0.025	0.029412	0.0375	0.013889	0.043478
15	0.041667	0.025	0.011765	0.05	0.055556	0.086957
16	0.402778	0.175	0.370588	0.3375	0.388889	0.282609
17	0.319444	0.55	0.341176	0.325	0.277778	0.326087
18	0.180556	0.15	0.194118	0.2	0.166667	0.152174
19	0.041667	0.05	0.041176	0.0375	0.083333	0.065217
20	0	0.025	0.011765	0	0.013889	0.043478
21	0	0	0	0.0125	0	0

Table 7. D16S539

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
8	0	0	0.005882	0	0	0
9	0.083333	0.025	0.082353	0.1	0.083333	0.086957
10	0.236111	0.275	0.182353	0.1625	0.166667	0.26087
11	0.25	0.3	0.241176	0.2875	0.236111	0.326087
12	0.375	0.325	0.441176	0.375	0.388889	0.23913
13	0.041667	0.075	0.047059	0.0625	0.111111	0.065217
14	0.013889	0	0	0.0125	0.013889	0.021739

Table 8. TPOX

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
8	0.291667	0.4	0.370588	0.35	0.347222	0.369565
9	0.041667	0	0.041176	0.0125	0.055556	0.043478
10	0.013889	0.05	0.047059	0.0125	0.055556	0.021739
11	0.486111	0.475	0.417647	0.4625	0.402778	0.347826
12	0.166667	0.075	0.123529	0.1625	0.138889	0.217391

Table 9. CSF1PO

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
8	0	0	0.005882	0	0	0
9	0.083333	0.25	0.152941	0.1	0.083333	0.130435
10	0.263889	0.275	0.2	0.25	0.333333	0.195652
11	0.166667	0.175	0.211765	0.225	0.125	0.26087
12	0.402778	0.25	0.382353	0.325	0.388889	0.304348
13	0.083333	0.05	0.041176	0.1	0.069444	0.108696
14	0	0	0.005882	0	0	0

Table 10. D8S1179

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
8	0	0	0.005882	0.0125	0	0
10	0.027778	0	0.029412	0.0125	0.055556	0.043478
11	0	0.025	0.011765	0.05	0.041667	0.065217
12	0.152778	0.175	0.1	0.1625	0.055556	0.130435
13	0.402778	0.35	0.382353	0.2875	0.472222	0.304348
14	0.333333	0.275	0.335294	0.375	0.291667	0.347826
15	0.041667	0.175	0.105882	0.0875	0.069444	0.108696
16	0.041667	0	0.029412	0.0125	0.013889	0

Table 11. D21S11

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
27	0.013889	0	0.011765	0.0125	0.027778	0
28	0.111111	0.025	0.041176	0.1125	0.041667	0.086957
29	0.152778	0.225	0.182353	0.1875	0.125	0.130435
30	0.194444	0.275	0.258824	0.2625	0.319444	0.326087
30.2	0.013889	0.05	0.011765	0.0125	0.027778	0.043478
31	0.027778	0.075	0.052941	0.05	0.027778	0.021739
31.2	0.222222	0.1	0.229412	0.1875	0.166667	0.173913
32	0	0	0.047059	0.0125	0.027778	0
32.2	0.208333	0.225	0.111765	0.1375	0.180556	0.108696
33	0	0.025	0	0	0	0
33.2	0.055556	0	0.052941	0.025	0.055556	0.108696

Table 12. D18S51

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
10	0	0	0.011765	0.0125	0	0.021739
11	0.027778	0	0.017647	0.0375	0.041667	0
12	0.083333	0.1	0.070588	0.1	0.055556	0.108696
13	0.069444	0.125	0.1	0.1125	0.027778	0.130435
14	0.347222	0.45	0.435294	0.35	0.402778	0.217391
15	0.138889	0.075	0.117647	0.075	0.222222	0.130435
16	0.069444	0.075	0.052941	0.0375	0.097222	0.130435
17	0.069444	0.025	0.076471	0.1375	0.027778	0.065217
18	0.013889	0.075	0.029412	0.0375	0.027778	0.043478
19	0.013889	0.025	0.023529	0.025	0.055556	0.043478
20	0	0	0	0	0	0.021739
21	0.027778	0	0.029412	0.0125	0.027778	0
22	0.069444	0.025	0.011765	0.05	0	0.065217
23	0.069444	0.025	0.017647	0.0125	0.013889	0.021739
24	0	0	0.005882	0	0	0

Table 13. D2S441

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
9	0.013889	0	0.011765	0	0	0
10	0.25	0.35	0.329412	0.3	0.208333	0.217391
11	0.611111	0.525	0.482353	0.575	0.527778	0.565217
11.3	0.027778	0.025	0.017647	0.0125	0	0
12	0.013889	0	0.029412	0.025	0.055556	0.021739
13	0	0	0.011765	0	0	0
13.1	0	0	0	0	0	0.021739
14	0.083333	0.1	0.111765	0.075	0.194444	0.152174
15	0	0	0.005882	0.0125	0.013889	0.021739

Table 14. TH01

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
5	0.013889	0	0.005882	0	0	0
6	0.125	0.075	0.088235	0.125	0.138889	0.108696
7	0.569444	0.65	0.652941	0.65	0.611111	0.586957
8	0.041667	0	0.029412	0.0375	0.013889	0.043478
9	0.013889	0.15	0.011765	0.0625	0.083333	0.086957
9.3	0.236111	0.1	0.205882	0.125	0.152778	0.173913
10	0	0.025	0.005882	0	0	0

Table 15. D19S433

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
12	0.013889	0.075	0.011765	0.025	0.027778	0.043478
12.2	0	0	0.005882	0.0125	0.013889	0
13	0.152778	0.05	0.111765	0.1	0.152778	0.130435
13.2	0.166667	0.3	0.182353	0.1625	0.305556	0.173913
14	0.263889	0.15	0.294118	0.2625	0.194444	0.195652
14.2	0	0.025	0.011765	0.0125	0	0
15	0.111111	0.2	0.111765	0.175	0.138889	0.130435
15.2	0.152778	0.1	0.117647	0.175	0.097222	0.130435
16	0.041667	0.05	0.064706	0.0375	0.013889	0.021739
16.2	0.069444	0.025	0.023529	0	0.027778	0.086957
17.2	0.013889	0.025	0.011765	0.0125	0.027778	0.086957
18.2	0.013889	0	0.052941	0.025	0	0

Table 16. FGA

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
18	0	0.025	0	0.025	0	0
18.2	0	0	0.005882	0	0	0
19	0.138889	0.075	0.129412	0.075	0.152778	0.043478
20	0.111111	0.075	0.064706	0.1	0.069444	0.086957
21	0.041667	0.1	0.076471	0.0875	0.055556	0.130435
22	0.083333	0.075	0.082353	0.1	0.194444	0.108696
22.2	0.013889	0	0	0	0	0
23	0.180556	0.275	0.2	0.1875	0.083333	0.173913
24	0.138889	0.175	0.164706	0.2375	0.305556	0.173913
25	0.180556	0.075	0.164706	0.1	0.097222	0.108696
26	0.097222	0.075	0.088235	0.0625	0.013889	0.130435
27	0	0.05	0.023529	0.025	0.013889	0.021739
28	0.013889	0	0	0	0.013889	0.021739

Table 17. D13S317

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
8	0.041667	0	0.058824	0.05	0.083333	0.086957
9	0.152778	0.175	0.2	0.2875	0.180556	0.130435
10	0.138889	0.3	0.141176	0.1	0.180556	0.217391
11	0.208333	0.125	0.135294	0.2	0.152778	0.086957
12	0.236111	0.175	0.311765	0.225	0.277778	0.369565
13	0.152778	0.15	0.111765	0.1	0.097222	0.108696
14	0.069444	0.075	0.041176	0.0375	0.027778	0

Table 18. D22S1045

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
11	0.083333	0.25	0.088235	0.1125	0.166667	0.065217
13	0	0	0.005882	0	0	0
14	0.013889	0	0.029412	0.025	0.041667	0.021739
15	0.527778	0.325	0.376471	0.3625	0.388889	0.326087
16	0.291667	0.35	0.4	0.425	0.361111	0.478261
17	0.027778	0.075	0.058824	0.05	0.027778	0.043478
18	0	0	0.011765	0.0125	0	0.043478
19	0.055556	0	0.029412	0.0125	0.013889	0.021739

Table 19. D5S818

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
7	0.291667	0.125	0.194118	0.1875	0.138889	0.282609
9	0.041667	0.075	0.070588	0.025	0.041667	0.021739
10	0.013889	0.075	0.058824	0.05	0	0.043478
11	0.416667	0.4	0.447059	0.5	0.555556	0.478261
12	0.166667	0.175	0.117647	0.125	0.194444	0.152174
13	0.069444	0.075	0.1	0.1125	0.069444	0
14	0	0.075	0.011765	0	0	0.021739

Table 20. D7S820

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
8	0.083333	0.05	0.141176	0.1375	0.083333	0.086957
9	0.083333	0.05	0.058824	0.0875	0.069444	0.065217
10	0.263889	0.275	0.282353	0.25	0.388889	0.304348
11	0.361111	0.475	0.347059	0.4	0.236111	0.369565
12	0.194444	0.15	0.158824	0.125	0.208333	0.173913
13	0.013889	0	0.011765	0	0.013889	0

Table 21. SE33

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
12	0.013889	0	0	0	0	0
14	0	0	0.005882	0.0125	0	0
15	0.013889	0.025	0.035294	0.075	0.083333	0.021739
16	0.097222	0.025	0.064706	0.0625	0.069444	0.130435
17	0.152778	0.125	0.088235	0.075	0.138889	0.043478
18	0.041667	0.025	0.052941	0.0875	0.055556	0.043478
18.2	0	0	0.005882	0	0	0
19	0.055556	0.05	0.094118	0.05	0.097222	0.086957
20	0.027778	0.05	0.023529	0.025	0.055556	0
21	0.013889	0.025	0.011765	0	0.027778	0.021739
21.2	0.027778	0	0.005882	0	0.041667	0.021739
22	0	0	0.011765	0	0	0
22.2	0.027778	0.025	0.005882	0.025	0.013889	0
23.2	0.013889	0	0.005882	0	0	0
24.2	0.013889	0.025	0	0	0	0
25.2	0.013889	0.025	0.005882	0.0125	0.013889	0.021739
26.2	0.041667	0.05	0.052941	0.0625	0.013889	0
27.2	0.125	0.15	0.188235	0.2	0.138889	0.23913
28.2	0.180556	0.15	0.105882	0.0875	0.083333	0.130435
29.2	0.013889	0.15	0.094118	0.1625	0.055556	0.086957
30.2	0.083333	0.05	0.1	0.0375	0.055556	0.130435
31.2	0.013889	0.05	0.023529	0.0125	0.041667	0.021739
32.2	0.013889	0	0.005882	0	0.013889	0
33	0.013889	0	0.005882	0.0125	0	0
34	0	0	0.005882	0	0	0

Table 22. D10S1248

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
10	0.013889	0	0.017647	0	0.013889	0
11	0.111111	0.075	0.070588	0.075	0.069444	0.043478
13	0.180556	0.15	0.205882	0.225	0.222222	0.217391
14	0.388889	0.575	0.464706	0.4125	0.361111	0.521739
15	0.194444	0.2	0.117647	0.125	0.194444	0.195652
16	0.055556	0	0.105882	0.1375	0.125	0
17	0.041667	0	0.017647	0.025	0.013889	0.021739
18	0.013889	0	0	0	0	0

Table 23. D1S1656

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
10	0	0	0.011765	0	0	0
11	0.013889	0	0.017647	0	0.013889	0.043478
12	0.055556	0	0.052941	0.05	0.055556	0.043478
13	0.111111	0.075	0.117647	0.0625	0.055556	0.152174
14	0.111111	0.025	0.117647	0.075	0.069444	0.086957
14.3	0	0	0.005882	0	0	0
15	0.208333	0.175	0.158824	0.075	0.152778	0.108696
15.3	0.027778	0	0.005882	0	0.027778	0.043478
16	0.194444	0.375	0.211765	0.1875	0.208333	0.26087
16.3	0.027778	0.025	0.035294	0.0875	0.055556	0.021739
17	0.027778	0.025	0.041176	0.0625	0.027778	0.043478
17.3	0.138889	0.2	0.152941	0.325	0.166667	0.086957
18	0.013889	0	0	0	0	0
18.3	0.069444	0.1	0.070588	0.075	0.166667	0.108696

Table 24. D12S391

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
15	0.027778	0	0.017647	0.0125	0.013889	0.021739
17	0.111111	0.05	0.041176	0.05	0.069444	0.108696
17.3	0.013889	0	0	0	0	0.021739
18	0.180556	0.15	0.211765	0.225	0.208333	0.108696
18.3	0	0	0	0	0.013889	0
19	0.166667	0.25	0.264706	0.15	0.180556	0.195652
19.2	0.069444	0	0.041176	0.0375	0.055556	0
20	0.083333	0.3	0.158824	0.175	0.180556	0.26087
21	0.069444	0.025	0.029412	0.1125	0.041667	0.021739
22	0.041667	0.025	0.041176	0.075	0.097222	0.043478
23	0.222222	0.15	0.135294	0.125	0.097222	0.152174
24	0.013889	0.05	0.035294	0.025	0.027778	0.021739
25	0	0	0.011765	0.0125	0.013889	0.043478
26	0	0	0.011765	0	0	0

Table 25. D2S1338

Allele	*CRSR*	*LBCC*	*PR*	*R*	*W*	*Y*
16	0.013889	0	0.005882	0.025	0.041667	0.021739
17	0.069444	0.025	0.076471	0.0625	0.069444	0.065217
18	0.041667	0.075	0.041176	0.0375	0.027778	0.065217
19	0.236111	0.375	0.205882	0.2625	0.180556	0.217391
20	0.069444	0.025	0.070588	0.125	0.069444	0.021739
21	0.027778	0.05	0	0.025	0.013889	0.043478
22	0.263889	0.225	0.3	0.2625	0.319444	0.173913
23	0.194444	0.175	0.188235	0.1	0.180556	0.282609
24	0.069444	0.025	0.052941	0.0875	0.083333	0.043478
25	0.013889	0.025	0.058824	0	0.013889	0.065217
26	0	0	0	0.0125	0	0

Table 26. Random Match Probability Calculation Comparisons

Locus	*A*	*AA*	*C*	*H*	*SI*
D3S1358	0.134965	0.10651	0.076628	0.094234	0.1874367
vWA	0.075897	0.061406	0.063908	0.071045	0.12178433
D16S539	0.087279	0.070453	0.094829	0.074244	0.113288
CSF1PO	0.098355	0.083052	0.129124	0.127084	0.100237
TPOX	0.22325	0.089728	0.178314	0.147558	0.174182
D8S1179	0.040509	0.065328	0.060723	0.061143	0.11715855
D21S11	0.056553	0.039844	0.049708	0.04481	0.0503639
D18S51	0.045613	0.026918	0.028398	0.026611	0.05256026
D2S441	0.102685	0.093313	0.087186	0.103525	0.20046547
D19S433	0.067601	0.036009	0.084378	0.066078	0.04318284
TH01	0.140465	0.101914	0.091882	0.085332	0.2255171
FGA	0.0403	0.028958	0.035636	0.026116	0.031416
D22S1045	0.096703	0.053925	0.121506	0.04881	0.156891
D5S818	0.072054	0.100762	0.143717	0.127054	0.119835
D13S317	0.068596	0.133943	0.076674	0.049521	0.057201
D7S820	0.099219	0.086171	0.059672	0.080771	0.105404
SE33	0.007054	0.008918	0.005164	0.006805	0.01453861
D10S1248	0.086323	0.066395	0.09791	0.100995	0.11220172
D1S1656	0.042387	0.031003	0.018853	0.024193	0.03291899
D12S391	0.044372	0.034762	0.021913	0.026759	0.03693331
D2S1338	0.030661	0.021188	0.02568	0.028526	0.053249
Overall	1.34231E-25	2.02306E-27	9.46137E-27	3.2616E-27	1.0281E-23

A – Asian Population, AA- African American Population, C- Caucasian Population, H-Hispanic Population

Table 27. Probability of Exclusion with Mother Child Data Calculation Comparisons

Locus	*A*	*AA*	*C*	*H*	*SI*
D3S1358	0.461331	0.516281	0.58467	0.542725	0.388548
vWA	0.587189	0.628546	0.620605	0.600913	0.488334
D16S539	0.557997	0.602227	0.543558	0.591696	0.502689
CSF1PO	0.535075	0.571659	0.470841	0.475618	0.528598
TPOX	0.342842	0.554323	0.398767	0.445219	0.401203
D8S1179	0.697601	0.617646	0.630207	0.630307	0.499313
D21S11	0.643977	0.701654	0.666052	0.682881	0.663907
D18S51	0.680696	0.754885	0.748064	0.75629	0.653932
D2S441	0.525912	0.547941	0.561621	0.524998	0.369433
D19S433	0.613451	0.715955	0.569827	0.61721	0.68882
TH01	0.453505	0.525494	0.547633	0.562445	0.347653
FGA	0.699377	0.745684	0.71732	0.758544	0.734523
D22S1045	0.534989	0.652138	0.489952	0.438769	0.431426
D5S818	0.597344	0.530135	0.450304	0.482527	0.493104
D13S317	0.6068	0.468581	0.587737	0.665747	0.640757
D7S820	0.532284	0.561907	0.633024	0.576124	0.518999
SE33	0.87672	0.860788	0.894944	0.87896	0.820835
D10S1248	0.561894	0.614322	0.534588	0.52861	0.506759
D1S1656	0.691552	0.735834	0.795658	0.767931	0.728412
D12S391	0.684884	0.720655	0.779396	0.755855	0.712611
D2S1338	0.73822	0.783015	0.760558	0.747432	0.655346
Overall	0.999999999	0.99999999981	0.99999999962	0.9999999996	0.99999999

A – Asian Population, AA- African American Population, C- Caucasian Population, H-Hispanic Population

Table 28. Probability of Exclusion with Child Data Calculation Comparisons

Locus	*A*	*AA*	*C*	*H*	*SI*
D3S1358	0.291765	0.340203	0.405907	0.364333	0.221471
vWA	0.408144	0.452839	0.443553	0.423049	0.315418
D16S539	0.377532	0.423831	0.365705	0.41313	0.32616
CSF1PO	0.355606	0.394346	0.299794	0.303988	0.349973
TPOX	0.192925	0.374527	0.231345	0.270802	0.242886
D8S1179	0.531313	0.441122	0.453465	0.454981	0.324476
D21S11	0.469557	0.538276	0.495253	0.515046	0.492873
D18S51	0.512977	0.604343	0.595134	0.60592	0.478332
D2S441	0.349254	0.369508	0.383356	0.348931	0.211579
D19S433	0.437339	0.555535	0.391197	0.440275	0.522167
TH01	0.281212	0.344061	0.368263	0.383432	0.179266
FGA	0.534821	0.592744	0.556418	0.608827	0.577713
D22S1045	0.357237	0.47934	0.315754	0.272816	0.267678
D5S818	0.419245	0.352407	0.282455	0.309762	0.311353
D13S317	0.428859	0.296618	0.409255	0.494026	0.465654
D7S820	0.353854	0.382735	0.457272	0.39744	0.341674
SE33	0.780483	0.756329	0.810113	0.784461	0.696181
D10S1248	0.383258	0.437604	0.357146	0.351519	0.325308
D1S1656	0.52552	0.580362	0.659472	0.621795	0.570599
D12S391	0.517845	0.561202	0.637289	0.606409	0.551199
D2S1338	0.582876	0.641538	0.611728	0.594644	0.483617
Overall	0.99999642	0.999999203	0.999998778	0.99999871	0.999985075

A – Asian Population, AA- African American Population, C- Caucasian Population, H-Hispanic Population

REFERENCES

1. Budowle B, Shea B, Niezgoda S, and Chakraborty R. CODIS STR Loci Data for 41 Sample Populations. J. Forensic. Sci. 2001; 46:453-489.
2. Butler JM. Forensic DNA Typing-Biology, Technology, and Genetics of STR Markers. 2005; 2nd ed. Elsevier Academic Press, Amsterdam.
3. Kanthaswamy S, Smith D.G. Genetic and Ethnohistoric Evidence Suggest Current Native American Population Datasets in the FBI's CODIS Database Are Not Sufficiently Representative. Forensic Sci. Int. Genet. (2014), http://dx.doi.org/ 10.1016/j.fsigen.2014.05.006
4. Budowle B, Chidambaram A, Strickland L, Beheim C, Taft G, Chakraborty R. Population Studies on Three Native Alaska Population Groups using STR Loci. Forensic Sci Int. 2002; 129:51-57.
5. Sun G. McGarvey S.T, Bayoumi R, Mulligan C. J, Barrantes R, Raskin S, Akey J, Chakraborty R, and Deka R. Global Genetic Variation at Nine Short Tandem Repeat Loci and Their Implications on Forensic Genetics. Eur. Jour. Hum. Genet. 2003; 11: 39-49.
6. Leffell M, Fallin M, Hildebrand W, Cavett J, Iglehart B, Zachary A. HLA Alleles and Haplotypes Among the Lakota Sioux: Report of the ASHI Minority Workshops, Part III. Hum Immunol. 2004; 65:78-89.
7. Smith S. The Y-haplotypes of the South Dakota Native American Sioux [dissertation]. Orlando, FL: University of Central Florida; 2003.

8. Lewis P. O., and Zaykin D. Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). 2001. Free program distributed by the authors over the internet from http://lewis.eeb.uconn.edu/lewishome/software.html
9. Chakraborty R, Stivers D. Paternity Exclusion by DNA Markers: Effects of Paternal Mutations. Journal of Forensic Sciences. 1996; 4: 671-677.
10. Chakraborty R, Smouse P, Neel J. Population Amalgamation and Genetic Variation: Observations on Artificially Agglomerated Tribal Populations of Central and South America. Am J Hum Genet. 1988; 43: 709-725.
11. Chakraborty R, Stivers D, Su B, Zhong Y, Budowle B. The Utility of Short Tandem Repeat Loci Beyond Human Identification: Implications for Development of New DNA Typing Systems. Electrophoresis. 1999; 20: 1682-1696.
12. National Research Council Committee on DNA Forensic Science, NRC . An Update: The Evaluation of Forensic DNA Evidence. 1996. National Academy Press, Washington D.C., Recommendation 4.1.
13. Hares D. Expanding the CODIS Core Loci in the United States. FSI Genetics. 2012; 6: 52-54.