Legal Medicine 23 (2016) 49–54

Contents lists available at ScienceDirect

Legal Medicine

journal homepage: www.elsevier.com/locate/legalmed

The genetic structure of native Americans in North America based on the Globalfiler® STRs

Kelly L. McCulloh [a], Jillian Ng [b], Robert F. Oldt [b], Jessica A. Weise [a], Joy Viray [c], Bruce Budowle [d,e], David Glenn Smith [a,b], Sreetharan Kanthaswamy [a,b,f,*]

[a] Forensic Science Graduate Program, University of California, One Shields Avenue, Davis, CA 95616, USA
[b] Molecular Anthropology Laboratory, Department of Anthropology, University of California, One Shields Avenue, Davis, CA 95616, USA
[c] Sacramento County District Attorney's Crime Laboratory, 4800 Broadway Suite 200, Sacramento, CA 95820, USA
[d] University of North Texas Health Science Center, 3500 Camp Bowie Boulevard, Fort Worth, TX 76107, USA
[e] Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia
[f] California National Primate Research Center, University of California, One Shields Avenue, Davis, CA 95616, USA

ARTICLE INFO

Article history:
Received 15 July 2016
Received in revised form 11 September 2016
Accepted 23 September 2016
Available online 23 September 2016

Keywords:
Forensic science
Native Americans
North America
Geography
Short tandem repeats
Population genetics

ABSTRACT

Current forensic STR databases, such as CODIS, lack population genetic data on Native American populations. Information from a geographically diverse array of tribes is necessary to provide improved statistical estimates of the strength of associations with DNA evidence. The Globalfiler® STR markers were used to characterize the genetic structure of ten tribal populations from seven geographic regions in North America, including those not presently represented in forensic databases. Samples from the Arctic region, Baja California, California/Great Basin, the Southeast, Mexico, the Midwest, and the Southwest were analyzed for allele frequencies, observed and expected heterozygosities, and F-statistics. The tribal samples exhibited an F_{ST} or θ value above the conservative 0.03 estimate recommended by the National Research Council (NRC) for calculating random match probabilities among Native Americans. The greater differentiation among tribal populations computed here (θ = 0.04) warrants the inclusion of additional regional Native American samples into STR databases.

1. Introduction

Population structure can be used to quantify genetic differentiation among subpopulations relative to the total population, and is expressed as F_{ST} [1] or theta (θ) [2]. F_{ST} determinations are necessary for calculating random match probabilities in forensic casework, as they provide investigators population genetic information to estimate match probabilities of a forensic sample to a known source. The National Research Council (NRC) [2] recommends that a correction factor value of F_{ST} or θ = 0.01 be used for general United States populations while a value of 0.03 be used for smaller and more isolated populations, such as Native Americans, where subdivision is more prevalent when determining genetic variation among populations.

Consistent with the NRC's recommendation, Budowle et al. [3] found that Native Americans exhibited the highest differentiation compared to Caucasian, Hispanic, African American, and Asian populations, with an F_{ST} estimate of 0.0282. While Caucasian Americans showed little or no genetic subdivision, the estimates of F_{ST} between Navajos and Apaches was 50 times that among African-Americans, 14 times that among Hispanic-Americans, and only 0.13 times of the value of the estimate for Asian-Americans [3]. This observation is especially significant because Navajo and Apache are closely related genetically. These tribes share a relatively recent common ancestry, which undoubtedly contributed to their F_{ST} value, even though both tribes have been highly admixed with different populations, including unrelated Native American tribes, for at least 500 years [4].

Furthermore, based on a study of 678 autosomal STR loci genotyped across 422 individuals from 29 Native American populations in North America, Central America, and South America [5], Native American tribes, including Chipewyan, Cree, Ojibwa (North America), Cabecar, Guaymi, Kaqchikel, Maya, Mixe, Mixtec, Pima, Zapotec (Central America), Arhuaco, Aymara, Embera, Huilliche, Inga, Kogi, Quechua, Waunana, Wayuu, Zenu (western South America), and Ache, Guarani, Kaingang, Karitiana, Piapoco, Surui, Ticuna [Arara], and Ticuna [Tarapaca] (eastern South America), showed greater differentiation than any other comparably sized

[*] Corresponding author at: School of Mathematics and Natural Sciences, ASU at the West Campus, 4701 W Thunderbird Rd., Glendale, AZ 85306-4908, USA.
E-mail address: sree.kanthaswamy@asu.edu (S. Kanthaswamy).

http://dx.doi.org/10.1016/j.legalmed.2016.09.007
1344-6223/© 2016 Elsevier Ireland Ltd. All rights reserved.

population (F_{ST} or $\theta = 0.08$). Therefore, the F_{ST} estimate from Wang et al. [5] suggests a higher F_{ST} than the 0.03 value currently recommended by the NRC [2] will be needed to adjust for population structure in forensic cases, including paternity testing, involving Native American individuals. To establish an informative Native American population database, a more detailed examination is necessary to determine whether significant differentiation exists to warrant the creation of additional Native American datasets. Given that the CODIS Native American STR database lacks tribes that are genetically similar to the vast majority of tribes living today and that geography is responsible for 60% of genetic differentiation [6], it is necessary to generate information for a more geographically diverse representation of additional tribes representing a greater number of geographic populations to better characterize genetic variation among Native Americans [4].

The current 13 CODIS loci are CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11 [7]. This study included eight additional autosomal loci (D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433, D22S1045, and SE33), which are included in the Globalfiler® PCR Amplification Kit (Applied Biosystems, Foster City, CA), and are included in the expanded CODIS core loci [8]. Profiling new Native American samples with these 21 loci will expand the existing pool of genetic profiles in the DNA database and provide more information on allele frequencies and population substructures. In addition, this study focused on the effect of geographic location on population structure and differentiation and quantified such variation. The STR typing of the geographically representative North American tribes from the Arctic region, Baja California, California/Great Basin, the Southeast, Mexico, the Midwest, and the Southwest establishes a more complete Native American database that can directly assist in forensic investigations as well as provide more reliable estimates of allele frequencies and genetic variation within and among the tribes.

2. Materials and methods

The Department of Anthropology Laboratory at UC Davis houses one of the largest databanks of geographically and linguistically representative full blood Native North American samples. Of the 3327 tribal DNA samples currently archived and available at the Department of Anthropology at UC Davis, the 418 samples from random individuals analyzed here were the only ones that met the quantification requirements for STR analysis. Prior approval from the UC Davis IRB (ID 430207-2) was obtained for the use of these samples for this study. The list of 418 tribal samples included in the study, as well as their geographic origins and mtDNA haplogroup distributions are shown in Table 1. In North America, haplogroup frequencies exhibit regional continuity that can be helpful in understanding relationships among the populations in those areas [9]. The geographical regions of the Native American tribes used in this study were based on Driver [10] and Lorenz and Smith [11]. Samples from the Southwest, Southeast, Midwest/Great Plains and Arctic region as well as samples from California/Great Basin, Baja California, and Mexico were included in this study.

2.1. Sample extraction

Samples consisting of serum, buffy coat, blood, or purified DNA were originally stored at −20 °C but have recently been maintained at 4 °C. DNA samples were extracted from serum, buffy coat, and blood samples using the QIAamp DNA Blood Mini Kit (QIAGEN, Redwood City, CA) following the manufacturer's protocol.

2.2. Sample quantification

DNA samples were quantified using the Quantifiler® Duo Quantification Kit and the 7500 Fast Real-time PCR system (Applied Biosystems). The quantification standards and DNA samples were both run in duplicate following the manufacturer's protocol.

2.3. Sample amplification

DNA samples were diluted to 1.0 ng/μL and amplified along with the National Institute of Standards and Technology (NIST) Standard Reference Material (SRM) 2391c reference DNA sample using the Globalfiler® PCR Amplification Kit (Applied Biosystems) according to the manufacturer's protocols. Amplified samples were diluted in Hi-Di Formamide (Applied Biosystems) and run on a 3130*xl* Genetic Analyzer with POP-4 polymer (Applied Biosystems) following manufacturer recommended parameters. The GeneScan™ 600 LIZ® Size Standard (Applied Biosystems) was used as the internal sizing standard and the Globalfiler® Allelic Ladder (Applied Biosystems) was used for sizing the alleles. Alleles were called using GeneMapperID-X v.1.4 (Applied Biosystems) with the Local Southern sizing method.

2.4. Statistical or other methods of data analyses

The extent of genetic variation within and among tribal samples, number of alleles, and observed and expected heterozygosity for each autosomal locus in each geographic region were calculated using Arlequin v3.5.1.2 [12]. Arlequin was also used to calculate the following F-statistics: F_{ST} – the proportion of genetic variance in a population that is due to differences among subdivisions within that population; F_{IS} – inbreeding coefficient, F_{IT}: total inbreeding coefficient, and pairwise F_{ST} – to assess the degree of differentiation between pairs of tribal samples which provides an insight into the historical connections among tribal samples and among the geographic regions these tribes represent. The statisti-

Table 1
The seven geographic samples represented by 10 tribes, their sample sizes (N), and mtDNA haplogroup frequencies. Tribes in the southwest US region of North America, such as Apache and Yavapai, have a high frequency of haplogroup B, a moderate frequency of haplogroup C, and low frequencies of haplogroups A, D, and X [11], while a few tribes in the northern half of Mexico, such as Huichol, and Cora, have lower frequencies of A, suggesting gene flow between the North American Southwest and Mexico [24].

Geographic region	Tribe	N	A	B	C	D	X	Refs.
Arctic	Eskimo	44	0.97	0	0	0.03	0	[11]
Baja CA	Cochimi	25	0.08	0.46	0.46	0	0	[11]
CA/Great Basin	Miwok	33	0.12	0.41	0.06	0.41	0	[11]
Southeast	Cherokee	34	0	0.31	0.31	0	0.38	[11]
Mexico	Cora	64	0.31	0.51	0.14	0.04	0	[24]
	Huichol	30	0.31	0.53	0.16	0	0	[24]
	Seri	29	0	0.13	0.86	0	0	[25]
Midwest	Chippewa	21	0.48	0.11	0.19	0	0.21	[11]
Southwest	Apache	88	0.62	0.17	0.14	0.07	0	[11]
	Yavapai	50	0	0.86	0.03	0.03	0.08	[11]

cal significance of the pairwise F_{ST} computations was determined with a probability distribution constructed from permutation tests (N = 1000) with Bonferroni corrections for multiple comparisons. Mann-Whitney *U* tests were performed to determine if population-specific estimates of diversity and F_{IS} differed significantly across populations and from the overall average. The Hardy-Weinberg Exact Test in the program GENEPOP 4.2 was used to determine if any of the tribal samples showed detectable deviations from expectations of equilibrium [13,14]. CONVERT v1.31 [15] was used to compute private allele frequencies (or alleles restricted to one group) at each locus within each geographically separate sample. Because differences in sample size can affect allele representation and estimates of genetic variation (particularly due to the presence or absence of rare alleles), each of the genetic parameters was recalculated using 1000 iterations of 21 randomly selected individuals from each tribe (Table 1) normalized to match that of the Chippewa tribe (N = 21).

3. Results

The Supplementary Table 1 presents allele frequencies across the 21 autosomal STR loci for each geographical region and the frequencies of the 10 individual tribes included in this study have been published in Ng et al. [16]. Table 2 presents the estimates of allele numbers (Na), and observed (OH) and expected (EH) heterozygosities across the geographic regions for all 21 STRs. The tribal samples averaged between 6 (Eskimo – Arctic) and 8 (Miwok – CA/Great Basin, Cherokee – Southeast, Cora – Mexico, and Apache (San Carlos Apache Reservation) and Yavapai – Southwest) alleles per locus. Estimates of allele numbers, both rare and common, based on 21 random individuals from each tribe suggest an influence of sample size; the difference between Na based on total sample and the sample of 21 is greatest for those tribes with the largest sample size (i.e., Cora – Mexico, and Apache and Yavapai – Southwest). The values of OH and EH in Table 2 did not appear to be influenced by sample size. OH values range from 0.68 (Eskimo – Arctic) to 0.78 (Miwok – CA/Great Basin) while EH values range from 0.69 (Eskimo – Arctic) to 0.77 (Cherokee – Southeast). Several private alleles among the tribes were identified with the Cherokee (Southeast) sample having the most (10), followed by Chippewa (Midwest – 5), Apache (Southwest – 5), Cora (Mexico – 5), Miwok (CA/Great Basin – 5), Yavapai (Southwest – 4), Huichol (Mexico – 1), and Seri (Mexico – 1) (Table 3). Frequencies of private alleles ranged from 0.006 to 0.005 (Table 3).

Table 3
Private alleles observed in this study: Midwest (5), CA/Great Basin (5), Mexico (7), Southwest (9), and Southeast (10).

Locus	Size	Tribe (Geographic region)	Frequency
vWA	21	Chippewa (Midwest)	0.024
CSF1PO	12.1	Apache (Southwest)	0.006
TPOX	6	Cherokee (Southeast)	0.030
TPOX	7	Cherokee (Southeast)	0.015
D21S11	24.2	Miwok (CA/Great Basin)	0.015
D21S11	27	Miwok (CA/Great Basin)	0.046
D21S11	29.2	Apache (Southwest)	0.011
D21S11	35.2	Yavapai (Southwest)	0.010
D18S51	9	Cherokee (Southeast)	0.030
D18S51	10	Cherokee (Southeast)	0.015
D18S51	11.2	Apache (Southwest)	0.006
D18S51	13.2	Cherokee (Southeast)	0.015
D18S51	23	Huichol (Mexico)	0.017
D2S441	12.3	Cherokee (Southeast)	0.015
D19S433	11	Yavapai (Southwest)	0.010
D19S433	17	Miwok (CA/Great Basin)	0.015
TH01	10.3	Cora (Mexico)	0.016
FGA	17	Cora (Mexico)	0.008
FGA	22.2	Miwok (CA/Great Basin)	0.030
FGA	26.2	Chippewa (Midwest)	0.024
FGA	29	Cora (Mexico)	0.008
D22S1045	10	Cherokee (Southeast)	0.015
D22S1045	12	Chippewa (Midwest)	0.024
D7S820	15	Cherokee (Southeast)	0.015
SE33	11	Cora (Mexico)	0.008
SE33	12	Yavapai (Southwest)	0.010
SE33	13.2	Yavapai (Southwest)	0.020
SE33	15.2	Apache (Southwest)	0.017
SE33	24	Cherokee (Southeast)	0.016
SE33	30	Apache (Southwest)	0.006
D10S1248	10	Chippewa (Midwest)	0.024
D1S1656	10	Cherokee (Southeast)	0.030
D1S1656	14.3	Chippewa (Midwest)	0.024
D1S1656	19	Seri (Mexico)	0.035
D12S391	17.3	Miwok (CA/Great Basin)	0.046
D12S391	19.3	Cora (Mexico)	0.008

Table 2
Allele number (Na), observed (OH) and expected (EH) heterozygosities for each tribe and geographic sample. Estimates based on 21 randomly chosen samples parenthesized show that sample size has not affected the analyses significantly. *Indicates tribal populations that conformed with HWE at $p < 0.01$ when all samples were included in the analyses. None of these populations deviated from HWE at $p < 0.01$ when 21 random samples from each population were analyzed.

Geographic region	Tribe	N	Na	OH	EH
Arctic	Eskimo	44	6 (6)	0.68 (0.67)	0.69 (0.71)
Baja CA	Cochimi	25	7 (7)	0.75 (0.74)	0.75 (0.75)
CA/Great Basin	Miwok	33	8 (7)	0.78 (0.76)	0.76 (0.76)
Southeast	Cherokee	34	8 (8)	0.74 (0.75)	0.77 (0.77)
Mexico	Cora*	64	8 (6)	0.70 (0.68)	0.73 (0.72)
Mexico	Huichol*	30	6 (6)	0.70 (0.69)	0.70 (0.71)
Mexico	Seri*	29	6 (5)	0.66 (0.67)	0.64 (0.64)
Midwest	Chippewa*	21	7 (7)	0.77 (0.77)	0.76 (0.76)
Southwest	Apache	88	8 (6)	0.73 (0.69)	0.73 (0.72)
Southwest	Yavapai*	50	8 (7)	0.74 (0.72)	0.73 (0.71)
Average estimates		*41.8 (21)*	*7.2 (6.5)*	*0.73 (0.71)*	*0.73 (0.73)*

Pairwise F_{ST}, as well as population-specific F_{ST}, and average F_{IS} are shown in Table 4; all pairwise F_{ST} p-values were statistically significant at the 0.05 level. Pairwise F_{ST} values from Table 4 suggest that differentiation among Native American tribes ranged from 0.006 (between Apache and Yavapai – Southwest) to 0.113 (between Eskimo – Arctic and Seri – Mexico). In addition to exhibiting the greatest levels of differentiation with each other, the Eskimo (Arctic) and Seri (Mexico) populations also exhibited the greatest differences from most of the study samples, with mean pairwise F_{ST} values of 0.073 and 0.070, respectively. The Arctic sample also showed genetic differences from other geographic samples that were correlated with geographic distance. Differentiation within the Continental US did not appear to be correlated with their geographical distances. Within Mexico, the mean pairwise F_{ST} among the Cora, Huichol, and Seri was approximately 0.05 with Cora and Huichol exhibiting the least differences (0.02) and Seri appearing to be the most genetically isolated. When the Cochimi tribe from Baja California was compared with the other

Table 4
Pairwise and population specific F_{ST} and F_{IS} based on the 22 autosomal STR loci in the seven geographic samples. Estimates based on 21 randomly chosen samples are above the diagonal. The overall F-statistics for all populations are $F_{IS} = 0.006$ (0.014), $F_{ST} = 0.039$ (0.041), and $F_{IT} = 0.045$ (0.056), where parenthesized values are estimates based on the 21 random samples.

Tribe (Geographic Region)	Eskimo (Arctic)	Cochimi (Baja California)	Miwok (CA/ Great Basin)	Cherokee (Southeast)	Cora (Mexico)	Huichol (Mexico)	Seri (Mexico)	Chippewa (Midwest)	Apache (Southwest)	Yavapai (Southwest)	F_{ST}	F_{IS}
Eskimo		0.057	0.066	0.052	0.074	0.083	0.090	0.029	0.045	0.051	0.074 (0.061)	0.017 (0.07)
Cochimi	0.073		0.019	0.017	0.027	0.038	0.067	0.024	0.027	0.028	0.034 (0.034)	0.002 (0.02)
Miwok	0.076	0.018		0.015	0.036	0.039	0.087	0.026	0.034	0.051	0.040 (0.041)	−0.029 (0)
Cherokee	0.064	0.016	0.012		0.036	0.042	0.076	0.019	0.032	0.032	0.035 (0.036)	0.034 (0.03)
Cora	0.072	0.020	0.029	0.026		0.019	0.052	0.029	0.025	0.030	0.032 (0.036)	0.040 (0.04)
Huichol	0.101	0.038	0.046	0.043	0.020		0.068	0.040	0.036	0.046	0.048 (0.046)	0.008 (0.02)
Seri	0.113	0.068	0.087	0.079	0.050	0.067		0.055	0.050	0.052	0.070 (0.066)	−0.043 (−0.05)
Chippewa	0.046	0.022	0.026	0.018	0.022	0.039	0.061		0.018	0.021	0.029 (0.029)	−0.020 (−0.02)
Apache	0.061	0.023	0.029	0.026	0.022	0.036	0.057	0.016		0.012	0.031 (0.031)	0.006 (0.04)
Yavapai	0.058	0.022	0.037	0.027	0.023	0.044	0.052	0.014	0.006		0.032 (0.036)	−0.011 (−0.02)

samples from Mexico, a range of pairwise F_{ST} from 0.02 (Cochimi-Cora) to 0.068 (Cochimi-Seri) was observed. It appears that geographic and genetic distances between Mexico and the other study samples are correlated.

F_{IS} values (Table 4) were highest for the Cora tribe from Mexico ($F_{IS} = 0.04$), followed by the Cherokee tribe (Southeast) and Eskimo (Arctic) samples ($F_{IS} = 0.034$ and 0.017, respectively). The other tribes exhibited either low (nearing zero) levels of F_{IS} values or none at all (negative values).

4. Discussion

Larger sample sizes tended to be more optimal than smaller ones for finding the most alleles or for computing genetic diversity estimates; for instance the decline in Na when samples of size 21 were analyzed is greatest for the largest sample sizes (i.e., Cora – Mexico, and Apache and Yavapai – Southwest). The same average number of 8 alleles per locus was observed in this study as in the Budowle et al. studies [3,17] which also used Apache and Eskimo samples albeit with much greater sample numbers. In spite of having screened many more individuals from the Apache, Athabaskan, Inupiat, and Yupik tribes, i.e. at least twice as many used here, Budowle et al. [3] reported slightly lower OH (0.70) as well as EH (0.71) in the Apache tribe and comparable OH and EH estimates among the Alaskan tribes; OH = 0.70 and average EH = 0.71.

Private STR alleles with a maximum frequency of 5% have been estimated in the present study. While no private allele with a frequency above 0.13 has been found [18], with the exception of a nine repeat allele (9RA) in D9S1120 which occurs at a high average frequency of 0.36 among tribal samples [19–21], the determination of population specific private alleles in this study, ranging from 1 (in the Seri and Huichol tribes of Mexico, respectively) to 10 (in the Cherokee from the Southeast) could further assist forensic investigators given their potential to differentiate tribal samples and to find perpetrators of specific tribal origin.

The higher F_{ST} values of the Arctic region for average and across all pairwise comparisons reflect the population's relative geographic isolation from the other populations. A Mann-Whitney U treatment of the heterozygosity and F_{IS} estimates revealed significantly ($p < 0.05$) lower heterozygosity estimates of the Arctic population (OH = 0.68 and EH = 0.69) in relation to the average across all other populations (OH = 0.73 and EH = 0.73). The higher F_{IS} value as compared to the total population average can be attributed to a lack of migration and an increase of non-random mating that also stems from genetic isolation. The Arctic population's low nuclear genetic variation based on OH and EH estimates is consistent with the population's mtDNA variation, which is almost exclusively mtDNA haplogroup A (average haplogroup A frequency = 0.97) [22].

In contrast to the Arctic population, other Native American populations have a wider range of mtDNA haplogroups (predominantly A, B, C, and D) with a few tribes having higher frequencies of haplogroup X [9] and an average F_{ST} value of 0.05, which is higher than all other sample comparisons if the Arctic tribe was not included. In Mexico, the Seri, Cora, and Huichol tribes, especially the Seri who have a relatively high tribe-specific F_{ST} value 0.07, are more isolated from the rest of the Mexican tribes since they live in inaccessible places, preserve their customs, and only reproduce among themselves [9].

The lower differentiation (pairwise $F_{ST} = 0.02$) between Cochimi (Baja CA) and Miwok (CA/Great Basin) compared to the differentiation between the former and Mexico ($F_{ST} = 0.04$) is consistent with the theory that coastal migration brought populations to the Baja peninsula [23]. The pairwise F_{ST} values between Baja CA and the rest of the populations (mean pairwise $F_{ST} < 0.05$) also suggest that Baja CA is not significantly differentiated from the rest of North America. The Yuman-speaking tribes of Baja California (including Cochimi, as well as Cucupa, Kiliwa, Kumiai, and Pai Pai, which were not analyzed here) were moved to their current location from their homeland in Mexico Proper, and are closely related to the Yuman-speaking tribes of the American Southwest (e.g., Hualapai and Yavapai), which can explain the lack of differentiation among those regions.

The Southwest (Apache and Yavapai) exhibited the lowest amount of differentiation ($F_{ST} = 0.02$) with the Midwest (Chippewa), which suggests that a high rate of gene flow between the Southwest and Midwest populations existed historically. MtDNA haplogroup A-D and X frequencies observed in the Southwest, Mexico, and North America also are consistent with high levels of gene flow among those regions [24]. Although the Southwest was slightly differentiated from the CA/Great Basin and Baja CA (range $F_{ST} = 0.02–0.04$) in this study, mtDNA haplogroup B, which is predominant in the Southwest, was also prevalent in the CA/Great Basin and northern Mexico. Since mid-continental migration of the Midwest and Southeast populations occurred more recently than the Pacific coastal and coastal interior migrations [22,23], such as the Cochimi, Miwok, Huichol, and Seri, less differentiation is expected ($F_{ST} = 0.02$ vs. $F_{ST} = 0.06$). The Arctic had the least amount of differentiation from the Midwest and Southeast (pairwise $F_{ST} = 0.05$ and 0.06, respectively) compared to the other populations, suggesting those two populations were the last to diverge from the Arctic.

The Southeast population was least differentiated from the Midwest (pairwise $F_{ST} = 0.02$) and CA/Great Basin populations (pairwise $F_{ST} = 0.01$), suggesting a migration out of the Northwest rather than from the west, as Fladmark [23] proposed. The F_{IS} value for the Midwest was -0.02, indicating a lack of genetic isolation, possibly due to migration through the Midwest into the Southeast after the glacial recession. Migration through the Midwest would bring in excess gene flow and would increase the amount of heterozygosity seen in that population. Alongside Mexico, the Southeast exhibited a high F_{IS} value (0.03), suggesting that population migration ended once the Atlantic Ocean was reached.

The present study shows that Native Americans exhibit greater overall inter-population differentiation ($F_{ST} = 0.04$) than reported by Budowle et al. [3] as would be expected with increased sample populations that are geographically heterogeneous. Wang et al.'s [5] study based on STRs (albeit not the CODIS STRs) computed F_{ST} values for the Americas that far exceeded the value obtained herein, especially for Central and South American populations ($F_{ST} = 0.06$ to 0.15). These tribes were not considered in the present study. However, they also observed a value of F_{ST} of 0.03 among the North American tribes of Chipewyan, Cree, and Ojibwa. While the North American FST estimate reported by Wang et al. [5] is more consistent with that of Budowle et al. [3] than with the present study, the three tribes in their study were all derived from the same geographic region and belong to the same language group [5]. Had these previous studies included more regionally representative unrelated tribes, their F_{ST} estimates would be at least comparable if not greater than the estimates obtained in the present study. Therefore, the present study does not support the NRC's recommendations [2] for using a correction factor of F_{ST} or θ of only 0.03 for calculating match probabilities in small isolated populations, such as the Native Americans. In fact, the present results show that a more stringent value of at least 0.04 should be used.

Since the CODIS Native American STR database contains only tribes from the Arctic and Subarctic regions and does not include the vast majority of other geographically diverse tribes, it is necessary to expand the database to include more unique genetic populations. Groups isolated by geography, such as the Arctic Eskimo and the Seri from Mexico, had the highest differentiation, while groups that have recently migrated out of the Northwest report low F_{ST} values. Expanding the study to include samples from Central and South America may increase the F_{ST} estimate [5]. Accurate F_{ST} values can help forensic investigators obtain more precise random match probabilities or make inferences of ethnic orgin in casework samples.

Acknowledgements

The research described in this article has been reported in Kelly L. McCulloh's Masters of Science (Forensic Science) thesis, which will be submitted to the UC Davis Office of Graduate Studies. The authors wish to thank the two anonymous reviewers for their critical review of this manuscript that has resulted in its significant imporvement. This study was supported by a National Institute of Justice – United States grant 2014-DN-BX-K024 to SK and a research grant to KM from the UC Davis Forensic Science Graduate Program.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.legalmed.2016.09.007.

References

[1] S. Wright, The interpretation of population structure by F-statistics with special regard to systems of mating, Evolution 19 (3) (1965) 395–420.
[2] National Research Council, The Evaluation of Forensic DNA Evidence, The National Academies Press, Washington DC, 1996.
[3] B. Budowle, B. Shea, S. Niezgoda, R. Chakraborty, CODIS STR loci data from 41 sample populations, J. Forensic Sci. 46 (3) (2001) 453–489.
[4] S. Kanthaswamy, D.G. Smith, Genetic and ethnohistoric evidence suggest current Native American population datasets in the FBI's CODIS database are not sufficiently representative, Forensic Sci. Int. Genet. 13 (2014) e13–e15.
[5] S. Wang, C.M. Lewis Jr., M. Jakobsson, S. Ramachandran, N. Ray, G. Bedoya, W. Rojas, M.V. Parra, J.A. Molina, C. Gallo, G. Mazzotti, G. Poletti, K. Hill, A.M. Hurtado, D. Labuda, W. Klitz, R. Barrantes, M.C. Bortolini, F.M. Salzano, M.L. Petzl-Erler, L.T. Tsuneto, E. Llop, F. Rothhammer, L. Excoffier, M.W. Feldman, N. A. Rosenberg, A. Ruiz-Linares, Genetic variation and population structure in native Americans, PLoS Genet. 3 (11) (2007) e185.
[6] E. Eller, Population substructure and isolation by distance in three continental regions, Am. J. Phys. Anthropol. 108 (2) (1999) 147–159.
[7] B. Budowle, T. Moretti, S. Niezgoda, B. Brown, CODIS and PCR-based short tandem repeat loci: law enforcement tools, in: Proceedings of the Second European Symposium on Human Identification, Promega Corporation, Madison, WI, 1998, pp. 73–88.
[8] D.R. Hares, Selection and implementation of expanded CODIS core loci in the United States, Forensic Sci. Int. Genet. 17 (2015) 33–34.
[9] R.I. Penaloza-Espinosa, D. Arenas-Aranda, R.M. Cerda-Flores, L. Buentello-Malo, G. Gonzalez-Valencia, J. Torres, B. Alvarez, I. Mendoza, M. Flores, L. Sandoval, F. Loeza, I. Ramos, L. Munoz, F. Salamanca, Characterization of mtDNA haplogroups in 14 Mexican indigenous populations, Hum. Biol. 79 (3) (2007) 313–320.
[10] H.E. Driver, Indians of North America, second ed., University of Chicago Press, Chicago, 1969.
[11] J.G. Lorenz, D.G. Smith, Distribution of four founding mtDNA haplogroups among Native North Americans, Am. J. Phys. Anthropol. 101 (3) (1996) 307–323.
[12] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): an integrated software package for population genetics data analysis, Evol. Bioinf. Online 1 (1) (2005) 47–50.
[13] M. Raymond, F. Rousset, GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism, J. Heredity 86 (3) (1995) 248–249.
[14] F. Rousset, Genepop'007: a complete re-implementation of the genepop software for Windows and Linux, Mol. Ecol. Resour. 8 (1) (2008) 103–106.
[15] J.C. Glaubitz, Convert: a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages, Mol. Ecol. Notes 4 (2) (2004) 309–310.
[16] J. Ng, R.F. Oldt, K.L. McCulloh, J.A. Weise, J. Viray, B. Budowle, D.G. Smith, S. Kanthaswamy, Native American population data based on the Globalfiler® autosomal STR loci, Forensic Sci. Int. Genet. 24 (2016) e12–e13.
[17] B. Budowle, A. Chidambaram, L. Strickland, C.W. Beheim, G.M. Taft, R. Chakraborty, Population studies on three Native Alaska population groups using STR loci, Forensic Sci. Int. 129 (1) (2002) 51–57.
[18] C. Phillips, A. Rodriguez, A. Mosquera-Miguel, M. Fondevila, L. Porras-Hurtado, F. Rondon, A. Salas, A. Carracedo, M.V. Lareu, D9S1120, a simple STR with a common Native American-specific allele: forensic optimization, locus characterization and allele frequency studies, Forensic Sci. Int. Genet. 3 (1) (2008) 7–13.
[19] K.B. Schroeder, T.G. Schurr, J.C. Long, N.A. Rosenberg, M.H. Crawford, L.A. Tarskaia, L.P. Osipova, S.I. Zhadanov, D.G. Smith, A private allele ubiquitous in the Americas, Biol. Lett. 3 (2) (2007) 218–223.
[20] H. Rangel-Villalobos, V.M. Sanchez-Gutierrez, M. Botello-Ruiz, J. Salazar-Flores, G. Martinez-Cortes, J.F. Munoz-Valle, C. Phillips, Evaluation of forensic and anthropological potential of D9S1120 in Mestizos and Amerindian populations from Mexico, Croatian Med. J. 53 (5) (2012) 423–431.
[21] I. Yuasa, Y. Irizawa, H. Nishimukai, Y. Fukumori, K. Umetsu, N. Nakayashiki, N. Saitou, L. Henke, J. Henke, A hypervariable STR polymorphism in the complement factor I (CFI) gene: Asian-specific alleles, Int. J. Legal Med. 125 (1) (2011) 121–125.
[22] J.S. Aigner, Early holocene evidence for the aleut maritime adaptation, Arctic Anthropol. 13 (2) (1976) 32–45.

[23] K.R. Fladmark, Routes: alternate migration corridors for early man in North America, Am. Antiq. 44 (1) (1979) 55–69.

[24] M.H. Snow, K.R. Durand, D.G. Smith, Ancestral puebloan mtDNA in context of the greater southwest, J. Archaeol. Sci. 37 (7) (2010) 1635–1645.

[25] R.S. Malhi, H.M. Mortensen, J.A. Eshleman, B.M. Kemp, J.G. Lorenz, F.A. Kaestle, J.R. Johnson, C. Gorodezky, D.G. Smith, Native American mtDNA prehistory in the American Southwest, Am. J. Phys. Anthropol. 120 (2) (2003) 108–124.