Forensic Science International: Genetics 24 (2016) 18–23

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Genetic analysis of the Yavapai Native Americans from West-Central Arizona using the Illumina MiSeq FGx™ forensic genomics system

Frank R. Wendt[a,*], Jennifer D. Churchill[a], Nicole M.M. Novroski[a], Jonathan L. King[a], Jillian Ng[b], Robert F. Oldt[b], Kelly L. McCulloh[c], Jessica A. Weise[c], David Glenn Smith[b,c], Sreetharan Kanthaswamy[b,c], Bruce Budowle[a,d]

[a] Institute of Applied Genetics, Department of Molecular and Medical Genetics, University of North Texas Health Science Center, 3500 Camp Bowie Blvd. Fort Worth, TX 76107, USA
[b] Molecular Anthropology Laboratory, Department of Anthropology, University of California, One Shields Avenue, Davis, CA 95616, USA
[c] Graduate Group in Forensic Science, University of California, One Shields Avenue, Davis, CA 95616, USA
[d] Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 15 February 2016
Received in revised form 10 May 2016
Accepted 14 May 2016
Available online 17 May 2016

Keywords:
ForenSeq™ DNA signature prep kit
Native american population
Yavapai
Allele frequencies
Population genetics
FGx™ system

ABSTRACT

Forensically-relevant genetic markers were typed for sixty-two Yavapai Native Americans using the ForenSeq™ DNA Signature Prep Kit.These data are invaluable to the human identity community due to the greater genetic differentiation among Native American tribes than among other subdivisions within major populations of the United States. Autosomal, X-chromosomal, and Y-chromosomal short tandem repeat (STR) and identity-informative (iSNPs), ancestry-informative (aSNPs), and phenotype-informative (pSNPs) single nucleotide polymorphism (SNP) allele frequencies are reported. Sequence-based allelic variants were observed in 13 autosomal, 3 X, and 3 Y STRs. These observations increased observed and expected heterozygosities for autosomal STRs by 0.081 ± 0.068 and 0.073 ± 0.063, respectively, and decreased single-locus random match probabilities by 0.051 ± 0.043 for 13 autosomal STRs. The autosomal random match probabilities (RMPs) were 2.37 × 10–26 and 2.81 × 10–29 for length-based and sequence-based alleles, respectively. There were 22 and 25 unique Y-STR haplotypes among 26 males, generating haplotype diversities of 0.95 and 0.96, for length-based and sequencebased alleles, respectively. Of the 26 haplotypes generated, 17 were assigned to haplogroup Q, three to haplogroup R1b, two each to haplogroups E1b1b and L, and one each to haplogroups R1a and I1. Male and female sequence-based X-STR random match probabilities were 3.28 × 10–7 and 1.22 × 10–6, respectively. The average observed and expected heterozygosities for 94 iSNPs were 0.39 ± 0.12 and 0.39 ± 0.13, respectively, and the combined iSNP RMP was 1.08 × 10–32. The combined STR and iSNP RMPs were 2.55 × 10–58 and 3.02 × 10–61 for length-based and sequence-based STR alleles, respectively. Ancestry and phenotypic SNP information, performed using the ForenSeq™ Universal Analysis Software, predicted black hair, brown eyes, and some probability of East Asian ancestry for all but one sample that clustered between European and Admixed American ancestry on a principal components analysis. These data serve as the first population assessment using the ForenSeq™ panel and highlight the value of employing sequence-based alleles for forensic DNA typing to increase heterozygosity, which is beneficial for identity testing in populations with reduced genetic diversity.

1. Introduction

There is a paucity of population genetic data on Native American populations. Since genetic differentiation among Native American tribes is likely to exceed that among subdivisions within other major populations residing in the US, population genetic data for various applications including forensic identity testing would be invaluable. The Yavapai are a Native American tribe with a semi-nomadic and hunter-gatherer history and a geographic distribution that spanned the Verde Valley in west-central Arizona and continued westward to the Colorado River [1]. The tribe's current population has a size of approximately 901 individuals and occupies territory spanning nearly 2000 acres around the Verde River north of the Gila River [2]. Churchill et al. [3] described the performance of the ForenSeq™ DNA Signature Prep Kit for multiplexed amplification and subsequent massively parallel

* Corresponding author.
E-mail address: Frank.Wendt@my.unthsc.edu (F.R. Wendt).

http://dx.doi.org/10.1016/j.fsigen.2016.05.008
1872-4973/

sequencing of autosomal, X-chromosomal, and Y-chromosomal short tandem repeats (STRs) and identity-informative (iSNPs), ancestry-informative (aSNPs), and phenotype-informative (pSNPs) single nucleotide polymorphisms (SNPs) in the same library preparation procedure. Such a genetic marker kit enables simultaneous typing of a large number of forensically relevant markers. Herein, 62 Yavapai Native Americans were typed with the broad collection of target markers in the ForenSeq™ panel and STR and SNP allele frequencies and typical population statistics were generated.

2. Materials and methods

2.1. Samples

Permission from the University of California Davis (UC Davis) Internal Review Board (IRB) (ID 430207-2) was obtained for the use of these samples for this study, and subsequent analyses were performed in accordance with both UC Davis and University of North Texas Health Science Center IRB approval. DNA was extracted from serum, buffy coat, and blood samples obtained from 64 anonymized Yavapai individuals of self-declared ancestry using the QIAamp DNA Blood Mini Kit (QIAGEN; Redwood City, CA) following the manufacturer's protocol. These samples are a subset of those reported by Smith et al. [5] and Monroe et al. [6] that had sufficient DNA quantity to be provided for this study. The serum albumin phenotypes [5] and mitochondrial DNA haplogroups and hypervariable region (HVR I and II) sequences [6] of these samples have been previously reported. The quantity of DNA for these samples was determined using the Quantifiler® Duo Quantification Kit and the 7500 Fast Real-time PCR system (Applied Biosystems; Foster City, CA). The quantification standards and DNA samples were both run in duplicate following the manufacturer's protocol. DNA was diluted to a final volume of 12 μl at a concentration of 0.2 ng/μl.

2.2. Library preparation

Libraries were prepared using the ForenSeq™ DNA Signature Prep Kit (Illumina; San Diego, Ca., USA) [3,4]. Five μl of genomic DNA per sample (1 ng total) were used for library preparation. DNA Primer Mix B (included in the ForenSeq™ kit, Illumina) was used to amplify and tag 58 STRs (27 autosomal, 7 X-chromosomal, and 24 Y-chromosomal), 94 iSNPs, 56 aSNPs, and 22 pSNPs. All thermal cycling was performed on an Applied Biosystems® GeneAmp® PCR System 9700 (Thermo Fisher Scientific) for one cycle at 98 °C for three minutes, eight cycles of 96 °C for 45 s, 80 °C for 30 s, 54 °C for two minutes, and 68 °C for two minutes, ten cycles of 96 °C for 30 s and 68 °C for three minutes, one cycle of 68 °C for ten minutes, and a final hold at 10 °C overnight. The tagged DNA subsequently was amplified to add index sequences for PCR product multiplexing using the following thermal cycling conditions: one cycle of 98 °C for 30 s, 15 cycles of 98 °C for 20 s, 66 °C for 30 s, and 68 °C for 90 s, one cycle of 68 °C for ten minutes, and a final hold at 10 °C. Bead purification followed by two 80% ethanol washes produced completed libraries, which were normalized and pooled in equal volume amounts according to the manufacturer's recommendations. Thirty-two Yavapai samples, one positive control, and two negative amplification controls were pooled per run. Ten μl of pooled, normalized libraries were diluted into 591 μl of hybridization buffer.

2.3. Massively parallel sequencing and data analysis

Massively parallel sequencing was performed on the MiSeq FGx™ desktop sequencer (Illumina) with a MiSeq ForenSeq™ sequencing kit (351 × 31 bp) (Illumina) according to the manufacturer's recommended protocol [7]. Data analysis was performed using the ForenSeq™ Universal Analysis Software (UAS) and RStudio® [8]. Additional analyses were performed using the "ForenSeqRunStatistics" XML file, generated by the ForenSeq™ UAS, with an in-house Excel-based workbook. Genetic Data Analysis (GDA) [9] was used for calculating observed and expected heterozygosity, and testing for departures from Hardy-Weinberg Equilibrium (HWE) expectations and linkage disequilibrium (LD). STR allele frequencies were determined by the counting method. STR allele sequence is reported according to Parson et al. [10]. Random match probabilities (RMPs) were calculated for autosomal STRs and iSNPs on a per marker basis and combined under the assumption of independence [11]. Y-STR haplogroups were determined using Haplogroup Predictor (http://www.hprg.com/hapest5/). Principle component analysis of the aSNPs and pSNP phenotypic estimates were performed on the ForenSeq™ UAS.

3. Results and discussion

Sixty-two of the 64 Yavapai samples yielded positive results after sequencing for 230 markers. After application of 20X depth of coverage (DoC) and 0.20 allele coverage ratio (ACR) thresholds, the average DoC and ACR values were 1455X ± 1458 and 0.76 ± 0.092 for STRs and 1209 ± 983.9 and 0.83 ± 0.069 for SNPs (Supplemental Figs. 1–4). Fifty-seven of the 62 positive results produced STR profiles that were ≥90% complete (a total of 58 loci per full male profile and 34 loci per full female profile) and 53 of these produced full SNP profiles (a total of 172 loci per full profile). Positive and negative controls performed as expected. Observed STR and SNP allele frequencies are shown in Supplemental Tables 1 and 2, respectively.

Tests for HWE detected 8 STRs and 12 SNPs ($p < 0.05$) that deviated from expectations. After Bonferroni correction, significant departures from HWE were observed for only 1 pSNP, N29insA ($p < 0.00024$, Supplemental Table 3). Considering the panel as a whole, the observed number of significant deviations from HWE is less than expected by chance alone (~10), and the population can be considered to meet HWE expectations. Linkage disequilibrium (LD) was assessed for the 27 autosomal STRs and 94 iSNPs first as separate groups. Twenty pairwise comparisons between STR loci and 144 pairwise comparisons between iSNP loci demonstrated significant LD ($p < 0.00014$ and 0.000011, respectively). The Penta E locus accounts for 17 of the 20 significant LD observations for STRs. The Penta E locus approaches statistical significance for deviation from HWE ($p = 0.001563$). In addition, the observed and expected homozygosity for the Penta E locus are 0.35 and 0.13, respectively. No detectable significant LDs were observed between STR loci on the same chromosomal arm. There were 144 significant LDs for the pairwise comparisons of iSNPs. The iSNP, rs2269355, accounts for 76 of the 144 significant deviations from linkage equilibrium. This marker is approaching statistical significance for departure from HWE ($p = 0.000313$), which may contribute to its apparent linkage at so many other loci pairs. Also, the observed and expected homozygosity for the rs2269355 locus are 0.54 and 0.77, respectively. No detectable statistically significant LDs were observed between SNP pairs on the same chromosomal arm. Of the 121 markers (7260 pairwise comparisons of autosomal STRs and iSNPs), 116 locus pairs demonstrated statistically significant linkage equilibrium, after Bonferroni correction ($p < 0.0000069$). As expected, the loci Penta E and rs2269355 account for 35% of the LD observations. No detectable statistically significant LDs were observed between STR and SNP pairs on the same chromosomal arm. The observed LD may be due to population substructure within the Yavapai population possibly resulting from admixture with nearby Apache groups with whom they maintained a close

relationship. However, the number of observed LDs is less than expected due to chance alone (~18 STR pairs, ~219 SNP pairs, ~363 STR and SNP pairs).

3.1. STRs

Intra-allelic sequence variants were identified at 19 of 59 STR loci (13 autosomal STRs, 3 X-STRs, and 4 Y-STRs), some of which were not reported by other comprehensive STR sequence-based analyses such as Gettings et al., Churchill et al., Warshauer et al., STRBase, or D'Amato et al. [12,3,13–15] (Supplemental Table 2). Observed and expected heterozygosity and RMPs were assessed for both length-based and sequence-based alleles for the autosomal STRs (Table 1). A SNP (rs9546005) 1 nucleotide downstream of the D13S317 repeat motif, an insertion-deletion (INDEL) in the repeat region adjacent to the D18S51 motif, and a SNP (rs73801920) adjacent to the D5S818 motif were captured by the ForenSeq™ UAS. Their contributions to sequence-based allele designation were included. There were examples of intra-allelic sequence variation at 13 autosomal STRs, which increased their observed heterozygosity by 0.084 ± 0.071 and decreased their single-locus-RMP values by an average of 0.051 ± 0.043. The loci D2S441 and D4S2408 had the largest percent increase in heterozygosity. However, sequence variants did not increase the observed heterozygosity of the FGA and vWA loci but did increase their expected heterozygosity by 0.0029 and 0.012, respectively. Although nominal, that increase in number of observed alleles was not captured as the allelic variants occurred in samples that were already heterozygous or were homozygous and both alleles contained the same allelic variation. Seven of the 13 autosomal STRs (D12S391, D13S317, D21S11, D2S441, D4S2408, D8S1179, and D9S1122) displayed >5% increase in expected heterozygosity with sequence-based alleles compared with length-based alleles. The observed heterozygosity for these 7 loci increased by 0.11 ± 0.047 and their single-locus-RMP values decreased by an average of 0.080 ± 0.037. The expected heterozygosities for these loci based on sequence variants were less than those seen in African American and Caucasian populations, which is consistent with the reduced diversity in Native American populations [16–18]. RMP values for length-based and sequence-based alleles for the 27 autosomal STRs were 2.37×10^{-26} and 2.81×10^{-29}, respectively. While both values are exceedingly low, the loci that carry sequence-based alleles may be worth considering for future STR panels. Loci with sequence variants and comparable or greater diversity than STRs lacking sequence variation are particularly desirable, because, in principle, they should amplify more robustly. These loci capture similar diversity in a smaller allele spread, minimizing the occurrence of large size differences between alleles of heterozygotes. Consequently, these loci may demonstrate less preferential amplification of smaller alleles which may lead to drop out. Using length-based STR genotypes from three major U.S. populations (African American, Caucasian, and Hispanic; unpublished data), a multidimensional scaling plot was generated based on pairwise genetic distances (Supplemental Fig. 5). While distant from all three comparison populations, the Yavapai are closer to the Hispanic population.

The heterozygosity of one X-STR, DXS10103 (females only), increased from 0.53 to 0.56 due to sequence-based alleles. Motif variants also were identified at the DXS10135 and DXS10074 loci, but heterozygosity was not assessed because they were only identified in hemizygous male samples. RMP values were calculated for males and females separately, according to Becker et al. [19]. Use of sequence-based alleles increased the single locus RMP value by 0.018 ± 0.019 and 0.043 ± 0.053 for males and females, respectively. The male X-STR 7-locus RMPs were 2.10×10^{-6} and 1.22×10^{-6} for length and sequence-based alleles, respectively (Supplemental Table 4). After Bonferroni correction ($p = 0.00238$), no significant LD was observed for any pairwise comparison. Assuming independence, the female X-STR 7-locus RMPs were 1.37×10^{-6} and 3.28×10^{-7} (Supplemental Table 4). It

Table 1
Comparison of observed (Ho) and expected heterozygosity (He), and random match probabilities (RMP) calculated using length-based and sequence-based allele designation for 62 samples (N = 62).

Locus	Ho (N = 62)		He (N = 62)		Locus RMP (N = 62)		
	Length-Based	Sequence-Based	Length-Based	Sequence-Based	Length-Based	Sequence-Based	Difference
CSF1PO	0.709677	0.709677	0.666929	0.666929	0.159021205	0.159021205	0
D10S1248	0.677419	0.677419	0.667453	0.667453	0.172470344	0.172470344	0
D12S391	0.672131	0.854938	0.769408	0.889851	0.089346622	0.025822375	0.063524247
D13S317	0.806452	0.83871	0.766719	0.816549	0.094229542	0.058540722	0.03568882
D16S539	0.790323	0.790323	0.74941	0.74941	0.107207455	0.107207455	0
D17S1301	0.66129	0.66129	0.712956	0.712956	0.142513839	0.142513839	0
D18S51	0.903226	0.919355	0.84802	0.851954	0.043751856	0.041783506	0.00196835
D19S433	0.822581	0.822581	0.819434	0.819434	0.059369547	0.059369547	0
D1S1656	0.737705	0.737705	0.805175	0.805175	0.062910525	0.062910525	0
D20S482	0.66129	0.66129	0.597299	0.597299	0.215312942	0.215312942	0
D21S11	0.7	0.790323	0.716667	0.886441	0.112057031	0.029264236	0.082792795
D22S1045	0.42623	0.42623	0.542542	0.542542	0.239827747	0.239827747	0
D2S1338	0.790323	0.822581	0.795961	0.835563	0.073483964	0.048007918	0.025476047
D2S441	0.612903	0.758065	0.569368	0.717021	0.252312058	0.129302471	0.123009588
D3S1358	0.677419	0.693548	0.598741	0.641621	0.238443938	0.18543028	0.053013658
D4S2408	0.596774	0.806452	0.561238	0.738264	0.227753061	0.103627702	0.124125358
D5S818	0.548387	0.596774	0.605822	0.619329	0.19093392	0.174927025	0.016006894
D6S1043	0.935484	0.935484	0.858773	0.858773	0.037924836	0.037924836	0
D7S820	0.725806	0.725806	0.710464	0.710464	0.134629975	0.134629975	0
D8S1179	0.83871	0.887097	0.813402	0.880803	0.064173512	0.028168697	0.036004815
D9S1122	0.596774	0.66129	0.551796	0.654472	0.257189587	0.165042792	0.092146795
FGA	0.935484	0.935484	0.85156	0.854445	0.043211169	0.041642647	0.001568521
PentaD	0.779661	0.779661	0.821672	0.821672	0.058189712	0.058189712	0
PentaE	0.648148	0.648148	0.873486	0.873486	0.030598514	0.030598514	0
TH01	0.548387	0.548387	0.591529	0.591529	0.218889268	0.218889268	0
TPOX	0.580645	0.580645	0.607658	0.607658	0.229763234	0.229763234	0
vWA	0.822581	0.822581	0.743247	0.755311	0.113184334	0.101836392	0.011347942
CombinedRMP	–	–	–	–	**2.3669E-26**	**2.8097E-29**	–

Table 2
Distribution of haplogroup assignments for 26 male Yavapai using Haplotype Predictor.

Highest Probability Haplogroup Assignment	Number of Samples (N = 26)
Q	17
R1b	3
E1b1b	2
L	2
R1a	1
I1	1

should be noted that, while not observed in this dataset after Bonferroni correction, significant LDs have been reported with X-STRs [20,21]. The combined RMP for X-STR profiles may not be appropriate for casework analyses; instead calculations may have to take into account linkage of X-STRs.

Complete Y-STR profiles were obtained for only five of 26 male samples due to poor typing success of the DYS392 locus, which also was observed by Churchill et al. [3]. Exclusion of this locus produced full profiles in 20 of the 26 male samples. Y-STR haplogroups were determined using Haplogroup Predictor by assuming equal prior ancestry odds (Table 2). The majority of Y-STR profiles (17 of the 26) were assigned to haplogroup Q, a branch of which, Q1a2a1a1 (or M3), is the predominant Y-chromosome haplogroup of Native America to which it is exclusive. The remaining nine haplotypes were assigned to the R1b, E1b1b, L, R1a, and I1 haplogroups, which probably reflect non-Native American admixture. The Q-haplogroup assignment is consistent with previously reported Y-chromosomal haplogroup studies of other Native American groups, including the Apache, who have historically resided in close proximity to the Yavapai [22]. These studies observed a Q-haplogroup frequency from 0.8–1.0 in Arizona. Malhi et al. [23] determined the R1a and R1b haplogroups to have the highest frequencies in the far north of North America (Subarctic region; Canada and Alaska) with frequencies ranging from 0 to 0.2 in the region of Yavapai inhabitance. Lastly, the E1b1b, L, and I1 haplogroups have documented origins out of northern Africa, India, and northern Europe, respectively. Compared to length-based Y-STR population data from Davis et al. [24] for three major U.S. populations (African American, Caucasian, and Hispanic) and three native Alaskan populations (Athabaskan, Yupik, and Inupiat), the Yavapai population is genetically closer to Hispanic and Inupiat individuals (Supplemental Fig. 6). Additionally, Monroe et al. [6] identified mitochondrial DNA haplogroups A–D (in order of descending relative haplogroup frequency) in a study including Yavapai and Western Apache populations. The presence of these haplogroups is consistent with those of indigenous peoples of the Americas.

There was low success in typing the DYS392 locus; only 6/26 Yavapai samples yielded results at this marker. Yavapai samples. Excluding this locus, 22 unique haplotypes were detected among the 26 male samples based on allele length with a haplotype diversity of 0.95 and 25 unique haplotypes when sequence variation was considered, which increased the haplotype diversity to 0.96 (Supplemental Table 5). Since the ForenSeq™ Y-STR panel lacks six loci and three loci, respectively, included in the Yfiler® Plus and PowerPlex® Y23 panels, only a subset of the loci in the Y-STR Haplotype Reference Database (YHRD) could be searched (Supplemental Table 5). A search for these partial profiles using YHRD Release 49 produced zero observations in datasets of 8,148 and 26,869 profiles, respectively.

3.2. SNPs

Allele frequencies for 94 iSNPs (Supplemental Table 3) were used to calculate single-locus RMP values and an overall iSNP RMP (Table 3). The average single-locus RMP value was 0.471 ± 0.131 and ranged from 0.375 (rs10776839) to 0.909 (rs2056277). The locus rs2056277 was previously reported by Poulsen et al. [25] to

Table 3
Observed (Ho) and expected (He) heterozygosity and random match probabilities (RMP) values for 94 identity-informative (iSNP) SNPs.

iSNP	Ho (N = 62)	He (N = 62)	RMP (N = 62)	iSNP	Ho (N = 62)	He (N = 62)	RMP (N = 62)	iSNP	Ho (N = 62)	He (N = 62)	RMP (N = 62)
rs1005533	0.214398	0.241935	0.642504026	rs174570	0.176239	0.16129	0.69620615	rs6811238	0.49764	0.467742	0.378247718
rs10092491	0.503541	0.580645	0.375260552	rs1821380	0.500787	0.435484	0.376641772	rs6955448	0.430857	0.428571	0.419179123
rs1015250	0.324548	0.241935	0.511598097	rs1886510	0.485182	0.548387	0.384891491	rs7041158	0.416667	0.383333	0.429705584
rs1024116	0.49764	0.500000	0.378247718	rs1979255	0.49764	0.500000	0.383240972	rs717302	0.121689	0.129032	0.780440293
rs1028528	0.238526	0.274194	0.610766118	rs2040411	0.503541	0.483871	0.375260552	rs719366	0.338166	0.295082	0.49794599
rs1031825	0.435957	0.45614	0.41584242	rs2046361	0.382198	0.344262	0.457405612	rs722098	0.499344	0.419355	0.377374202
rs10488710	0.456727	0.370968	0.401785488	rs2056277	0.0476	0.048387	0.908911206	rs722290	0.428534	0.483871	0.420880927
rs10495407	0.440598	0.387097	0.412422674	rs2076848	0.201941	0.225806	0.659563101	rs727811	0.314713	0.290323	0.52183004
rs1058083	0.503934	0.532258	0.375065062	rs2107612	0.314713	0.322581	0.52183004	rs729172	0.499344	0.548387	0.377374202
rs10773760	0.491302	0.555556	0.381536866	rs2111980	0.434697	0.435484	0.416506188	rs733164	0.304616	0.241935	0.532631863
rs10776839	0.503934	0.467742	0.375065062	rs214955	0.456727	0.532258	0.401785488	rs735155	0.490952	0.419355	0.381757426
rs1109037	0.361264	0.403226	0.475921458	rs221956	0.415421	0.451613	0.430561947	rs737681	0.107396	0.112903	0.803963538
rs1294331	0.419455	0.360656	0.427571066	rs2269355	0.461579	0.225806	0.398735925	rs740598	0.461579	0.419355	0.398735925
rs12997453	0.419455	0.393443	0.427571066	rs2342747	0.473649	0.393443	0.391488029	rs740910	0.502885	0.532258	0.375587383
rs13182883	0.401259	0.483871	0.441587549	rs2399332	0.314713	0.354839	0.52183004	rs763869	0.324548	0.33871	0.511598097
rs13218440	0.50359	0.508197	0.375269178	rs251934	0.121689	0.129032	0.780440293	rs8078417	0.470496	0.516129	0.393312523
rs1335873	0.481904	0.435484	0.386715988	rs279844	0.503997	0.590164	0.375067213	rs826472	0.167927	0.150000	0.708541696
rs1336071	0.448179	0.366667	0.407407407	rs2830795	0.361264	0.274194	0.475921458	rs873196	0.09284	0.096774	0.828538144
rs1355366	0.121689	0.129032	0.780440293	rs2831700	0.500787	0.532258	0.376641772	rs876724	0.369788	0.258065	0.468208492
rs1357617	0.079258	0.081967	0.852053001	rs2920816	0.427296	0.529412	0.422315667	rs891700	0.495673	0.516129	0.379266281
rs1360288	0.314713	0.322581	0.52183004	rs321198	0.487255	0.483333	0.383826418	rs901398	0.503541	0.354839	0.375260552
rs1382387	0.078023	0.080645	0.854197913	rs338882	0.49764	0.532258	0.378247718	rs907100	0.440598	0.483871	0.412422674
rs1413212	0.33412	0.322581	0.501913194	rs354439	0.287359	0.278689	0.551832884	rs914165	0.481904	0.435484	0.386715988
rs1454361	0.488198	0.467742	0.383240972	rs3780962	0.369788	0.387097	0.468208492	rs917118	0.401259	0.483871	0.441587549
rs1463729	0.440598	0.451613	0.412422674	rs430046	0.488198	0.500000	0.383240972	rs938283	0.361264	0.435484	0.475921458
rs1490413	0.272751	0.290323	0.568694161	rs4364205	0.324548	0.33871	0.511598097	rs964681	0.428534	0.483871	0.420880927
rs1493232	0.485182	0.354839	0.384891491	rs445251	0.499344	0.387097	0.384891491	rs987640	0.488198	0.467742	0.383240972
rs1498553	0.503541	0.548387	0.375260552	rs4530059	0.386048	0.387097	0.454089295	rs9905977	0.481904	0.532258	0.386715988
rs1523537	0.488198	0.435484	0.383240972	rs4606077	0.401259	0.387097	0.441587549	rs993934	0.434697	0.564516	0.416506188
rs1528460	0.369788	0.387097	0.468208492	rs560681	0.434697	0.467742	0.416506188	rs9951171	0.493443	0.564516	0.380434458
rs159606	0.352478	0.322581	0.484096193	rs576261	0.470496	0.548387	0.393312523	**Combined RMP**	–	–	**1.0759-E-32**

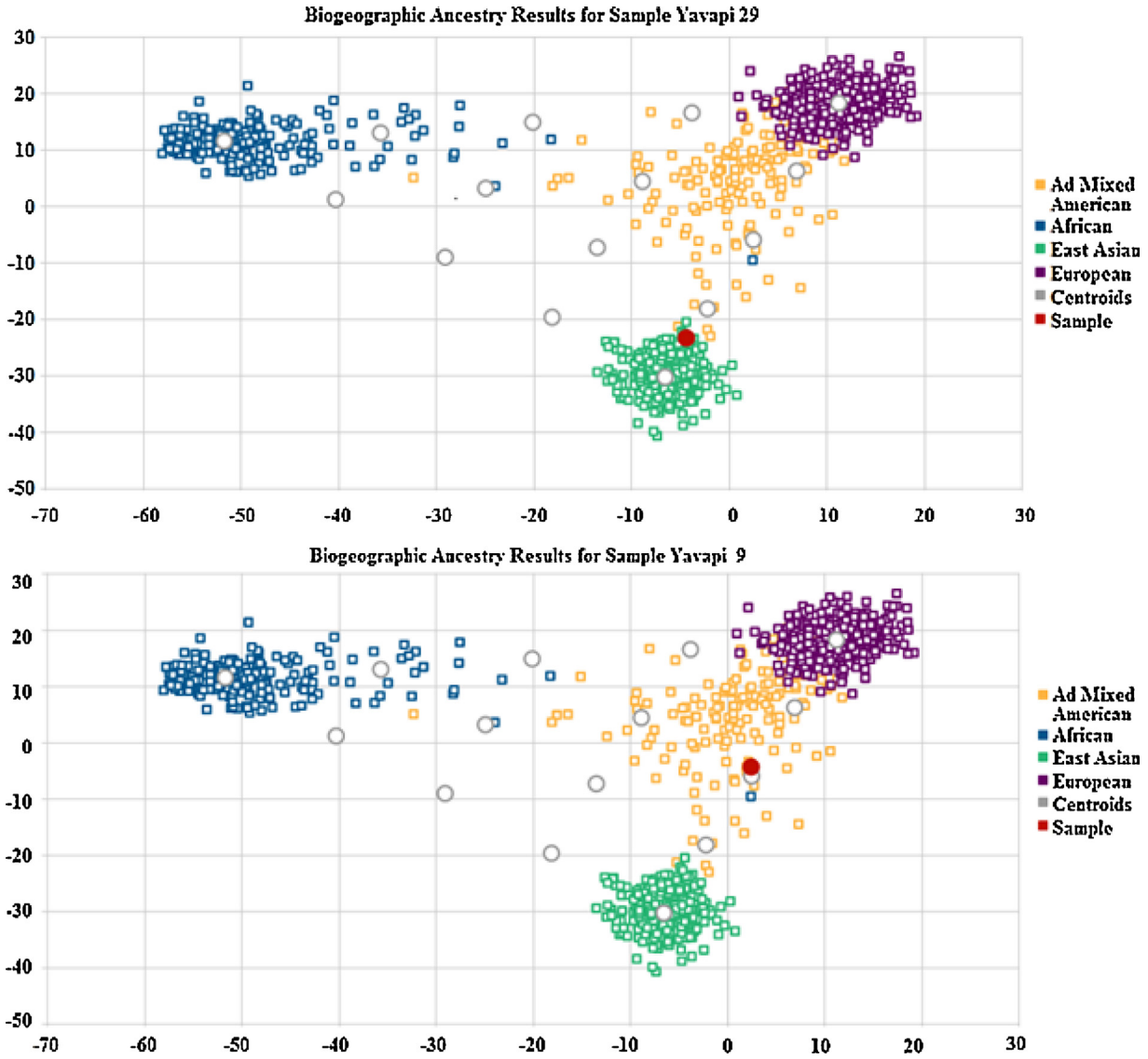

Fig. 1. Principal component analysis results for two Yavapai samples (#29 and #9) on the ForenSeq™ Universal Analysis Software. A red circle indicates the respective samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

be relatively uninformative in Native Americans for identity testing purposes. The resulting iSNP profile RMP was 1.08×10^{-32}.

Phenotypic and aSNP information were assessed using the ForenSeq™ UAS. Biogeographic ancestry analyses indicated that 21 samples clustered with East Asians, 22 individuals clustered with Admixed Americans, 18 individuals resided between East Asian and Admixed American, and one fell between Admixed American and European (Fig. 1). For example, Yavapai #73 fell between these populations. This is a male sample who was predicted with 100% probability to carry an R1b Y-haplogroup, the predominant Y-chromosome haplogroup among western Europeans. Concordance between autosomal aSNPs and the Y-chromosomal haplogroup information suggests that this observation is due to potential admixture. Full pSNP profiles were obtained for 59 samples, all of which, not surprisingly, were predicted to have black hair and brown eyes with average probabilities of 0.89 ± 0.086 and 0.99 ± 0.040, respectively.

4. Conclusion

While correcting for population substructure has been advocated for assigning random match probabilities, there is still a need to understand the genetic diversity of forensic genetic markers in various populations, particularly for more isolated populations such as Native Americans. This study serves as the first investigation into Yavapai Native American population genetics with respect to forensically-relevant loci as well as the first set of population data reported using the ForenSeq™ DNA Signature Prep Kit. STR and SNP allele frequencies have been reported for 62 Yavapai Native Americans of west-central Arizona. Random match probabilities were calculated producing values with much lower probabilities than those previously reported by Keisler et al. and Budowle et al. [26,27]. Twenty-five unique Y-STR haplotypes were observed (22 if only length-based alleles are considered); however, these data presently do not meet the minimum requirements for uploading to the YHRD database because the DYS392 and DYS393 loci are required for that purpose. The profiles will be uploaded

into the database upon completion of DYS392 and DYS393 typing by other means at a later date. While the Y-STR haplogroup predictions indicate consistent results with what is currently known about Native American origins, assessment of the mitochondrial DNA has provided insight into the maternal lineage of the Yavapai [6]. Biogeographic ancestry and phenotypic predictions indicated that the sample donors have black hair and brown eyes and generally cluster with East Asian and Admixed American populations. The data presented produce combined STR and SNP RMPs of 2.55×10^{-58} and 3.02×10^{-61} for length-based and sequence-based STR allele designation, respectively. The magnitude of these RMP values highlights the power of a combined STR and SNP approach towards source attribution in forensic DNA typing. These findings also demonstrate the value of increased heterozygosity and discrimination potential as a result of employing sequence-based STR allele designation.

Conflict of interest

The authors report no conflict of interest.

Acknowledgements

This study was partly supported by a National Institute of Justice grant 2014-DN-BX-K024 to SK and DGS and a research grant to KM from the UC Davis Forensic Science Graduate Program. Additionally, the authors would like to thank Illumina for kindly providing library preparation and sequencing reagents.

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.fsigen.2016.05.008.

References

[1] http://grandcanyonhistory.clas.asu.edu/history_nativecultures_yavapaiapache.html.
[2] http://arizonaexperience.org/people/indian-tribes-and-communities.
[3] J.D. Churchill, S.E. Schmedes, J.L. King, B. Budowle, Evaluation of the Illumina (®) beta version ForenSeq™ DNA signature prep kit for use in genetic profiling, Forensic. Sci. Int. Genet. 20 (2016) 20–29, doi:http://dx.doi.org/10.1016/j.fsigen.2015.09.009 (January, Epub 2015 Sep 21. PubMed PMID: 26433485).
[4] ForenSeq™ DNA Signature Prep Reference Guide. Document # 15049528. September 2015. http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/forenseq/forenseq-dna-signature-prep-guide-15049528-01.pdf.
[5] D.G. Smith, J. Lorenz, B.K. Rolfs, R.L. Bettinger, B. Green, J. Eshleman, B. Schultz, R. Malhi, Implications of the distribution of Albumin Naskapi and Albumin Mexico for new world prehistory, Am. J. Phys. Anthropol. 111 (April 4) (2000) 557–572 (PubMed PMID: 10727973).
[6] C. Monroe, B.M. Kemp, D.G. Smith, Exploring prehistory in the North American southwest with mitochondrial DNA diversity exhibited by Yumans and Athapaskans, Am. J. Phys. Anthropol. 150 (April 4) (2013) 618–631, doi:http://dx.doi.org/10.1002/ajpa.22237 (Epub 2013 Feb 26. PubMed PMID: 23440652).
[7] MiSeq System User Guide. https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-system-guide-15027617-o.pdf.
[8] RStudio Team 2015 RStudio: Integrated Development for R.R Studio, Inc., Boston MA URL http://www.rstudio.com/.
[9] Genetic Data Analysis Software. Lewis and Zaykin. 1999.
[10] Walther Parson, David Ballard, Bruce Budowle, John M. Butler, Katherine B. Gettings, Peter Gill, Leonor Gusmão, Douglas R. Hares, Jodi A. Irwin, Jonathan L. King, Peter de Knijff, Niels Morling, Mechthild Prinz, Peter M. Schneider, Christophe Van Neste, Sascha Willuweit, Christopher Phillips, Massively Parallel Sequencing of forensic STRs: Considerations of the DNA commission of the international society for forensic genetics (ISFG) on minimal nomenclature requirements, Forensic Science International: Genetics http://dx.doi.org/10.1016/j.fsigen.2016.01.009.
[11] National Research Council (U.S.). Committee on DNA Forensic Science. Evaluation of Forensic DNA Evidence. Washington, D.C.: National Academy, 1996. http://www.nap.edu/catalog5141.html.
[12] K.B. Gettings, R.A. Aponte, P.M. Vallone, J.M. Butler, STR allele sequence variation: current knowledge and future issues, Forensic. Sci. Int. Genet. 18 (2015) 118–130, doi:http://dx.doi.org/10.1016/j.fsigen.2015.06.005 (September, Epub 2015 Jul 6. PubMed PMID: 26197946.).
[13] D.H. Warshauer, J.D. Churchill, N. Novroski, J.L. King, B. Budowle, Novel Y-chromosome short tandem repeat variants detected through the use of massively parallel sequencing, Genomics Proteomics Bioinf. 13 (August 4) (2015) 250–257, doi:http://dx.doi.org/10.1016/j.gpb.2015.08.001 (Epub 2015 Sep 21. PubMed PMID: 26391384; PubMed Central PMCID: PMC4610967.).
[14] C.M. Ruitberg, D.J. Reeder, J.M. Butler, STRBase: a short tandem repeat DNA database for the human identity testing community, Nucleic. Acids Res. 29 (January 1) (2001) 320–322 (PubMed PMID: 11125125; PubMed Central PMCID: PMC29767.).
[15] M.E. D'Amato, L. Ehrenreich, K. Cloete, M. Benjeddou, S. Davison, Characterization of the highly discriminatory loci DYS449, DYS481, DYS518, DYS612, DYS626, DYS644 and DYS710, Forensic. Sci. Int. Genet. 4 (February 2) (2010) 104–110, doi:http://dx.doi.org/10.1016/j.fsigen.2009.06.011 (Epub 2009 Jul 29. PubMed PMID: 20129468.).
[16] M.D. Coble, C.R. Hill, J.M. Butler, Haplotype data for 23 Y-chromosome markers in four U.S. population groups, Forensic. Sci. Int. Genet. 7 (2013) e66–e68.
[17] C.R. Hill, D.L. Duewer, M.C. Kline, M.D. Coble, J.M. Butler, U.S. population data for 29 autosomal STR loci, Forensic. Sci. Int. Genet. 7 (2013) e82–e83.
[18] J.M. Butler, C.R. Hill, M.D. Coble, Variability of New STR Loci and Kits in U.S Population Groups. Profiles in DNA, (2012) (Available at http://www.promega.com/resources/articles/profiles-in-dna/2012/variability-of-new-str-loci-and-kits-in-us-population-groups/.).
[19] D. Becker, H. Rodig, C. Augustin, J. Edelmann, F. Götz, S. Hering, R. Szibor, W. Brabetz, Population genetic evaluation of eight X-chromosomal short tandem repeat loci using Mentype Argus X-8 PCR amplification kit, Forensic. Sci. Int. Genet. 2 (January 1) (2008) 69–74, doi:http://dx.doi.org/10.1016/j.fsigen.2007.08.013 (Epub 2007 Sep 29. PubMed PMID: 19083792).
[20] I. Gomes, M. Prinz, R. Pereira, C. Meyers, R.S. Mikulasovich, A. Amorim, A. Carracedo, L. Gusmão, Genetic analysis of three US population groups using an X-chromosomal STR decaplex, Int. J. Legal Med. 121 (May 3) (2007) 198–203 (Epub 2007 Jan 20. PubMed PMID: 17237948.).
[21] Q.L. Liu, J.Z. Wang, L. Quan, H. Zhao, Y.D. Wu, X.L. Huang, D.J. Lu, Allele and haplotype diversity of 26 X-STR loci in four nationality populations from China, PLoS One 8 (6) (2013) e65570 (June 21, Print 2013. PubMed PMID: 23805185; PubMed Central PMCID: PMC3689794.).
[22] M.C. Bortolini, F.M. Salzano, M.G. Thomas, S. Stuart, S.P. Nasanen, C.H. Bau, M.H. Hutz, Z. Layrisse, M.L. Petzl-Erler, L.T. Tsuneto, K. Hill, A.M. Hurtado, D. Castro-de-Guerra, M.M. Torres, H. Groot, R. Michalski, P. Nymadawa, G. Bedoya, N. Bradman, D. Labuda, A. Ruiz-Linares, Y-chromosome evidence for differing ancient demographic histories in the Americas, Am. J. Hum. Genet. 73 (July 3) (2003) 524–539 (Epub 2003 Jul 28. PubMed PMID: 12900798; PubMed Central PMCID: PMC1180678.).
[23] R.S. Malhi, A. Gonzalez-Oliver, K.B. Schroeder, B.M. Kemp, J.A. Greenberg, S.Z. Dobrowski, D.G. Smith, A. Resendez, T. Karafet, M. Hammer, S. Zegura, T. Brovko, Distribution of Y chromosomes among native North Americans: a study of Athapaskan population history, Am. J. Phys. Anthropol. 137 (December 4) (2008) 412–424, doi:http://dx.doi.org/10.1002/ajpa.20883 (PubMed PMID: 18618732; PubMed Central PMCID: PMC2584155.).
[24] C. Davis, J. Ge, A. Chidambaram, J. King, M. Turnbough, M. Collins, O. Dym, R. Chakraborty, A.J. Eisenberg, B. Budowle, Y-STR loci diversity in native Alaskan populations, Int. J. Legal. Med. 125 (July 4) (2011) 559–563, doi:http://dx.doi.org/10.1007/s00414-011-0568-3 (Epub 2011 Mar 30. PubMed PMID: 21448665.).
[25] L. Poulsen, C. Børsting, C. Tomas, F. González-Andrade, R. Lopez-Pulles, J. González-Solórzano, N. Morling, Typing of Amerindian kichwas and mestizos from Ecuador with the SNPforID multiplex, Forensic. Sci. Int. Genet. 5 (August 4) (2011) e105–7, doi:http://dx.doi.org/10.1016/j.fsigen.2011.03.006 (Epub 2011 Apr 9. PubMed PMID: 21481663.).
[26] K.M. Kiesler, P.M. Vallone, Allele frequencies for 40 autosomal SNP loci typed for US population samples using electrospray ionization mass spectrometry, Croat. Med. J. 54 (3) (2013) 225–231, doi:http://dx.doi.org/10.3325/cmj.2013.54.225.
[27] B. Budowle, R. Chakraborty, G. Carmody, K. Monson, Source attribution of a forensic DNA profile, Forensic. Sci. Commun. (2000).