Forensic Science International: Genetics 28 (2017) 146–154

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Short communication

Flanking region variation of ForenSeq™ DNA Signature Prep Kit STR and SNP loci in Yavapai Native Americans

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

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

ARTICLE INFO

Article history:
Received 16 December 2016
Received in revised form 1 February 2017
Accepted 24 February 2017
Available online 27 February 2017

Keywords:
Sequence variation
STR
SNP
Flanking region
ForenSeq™ DNA Signature Prep Kit
Massively parallel sequencing
STRait Razor v2s

ABSTRACT

Massively parallel sequencing (MPS) offers advantages over current capillary electrophoresis-based analysis of short tandem repeat (STR) loci for human identification testing. In particular STR repeat motif sequence information can be obtained, thereby increasing the discrimination power of some loci. While sequence variation within the repeat region is observed relatively frequently in some of the commonly used STRs, there is an additional degree of variation found in the flanking regions adjacent to the repeat motif. Repeat motif and flanking region sequence variation have been described for major population groups, however, not for more isolated populations. Flanking region sequence variation in STR and single nucleotide polymorphism (SNP) loci in the Yavapai population was analyzed using the ForenSeq™ DNA Signature Prep Kit and STRait Razor v2s. Seven and 14 autosomal STRs and identity-informative single nucleotide polymorphisms (iiSNPs), respectively, had some degree of flanking region variation. Three and four of these identity-informative loci, respectively, showed ≥5% increase in expected heterozygosity. The combined length- and sequence-based random match probabilities (RMPs) for 27 autosomal STRs were 6.11×10^{-26} and 2.79×10^{-29}, respectively. When combined with 94 iiSNPs (a subset of which became microhaplotypes) the combined RMP was 5.49×10^{-63}. Analysis of length-based and sequence-based autosomal STRs in STRUCTURE indicated that the Yavapai are most similar to the Hispanic population. While producing minimal increase in X- and Y-STR discrimination potential, access to flanking region data enabled identification of one novel X-STR and three Y-STR alleles relative to previous reports. Five ancestry-informative SNPs (aiSNPs) and two phenotype-informative SNPs (piSNPs) exhibited notable flanking region variation.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Forensic DNA typing currently utilizes length-based separation of polymerase chain reaction (PCR)-amplified short tandem repeats (STRs) for routine casework. Massively parallel sequencing (MPS) elucidates an additional level of STR motif variation at the sequence level, increasing diversity [1–3]. Sequence-based STR information may add value in kinship analyses and complex mixture de-convolution efforts. While some commonly used STR loci lack STR-motif sequence variation, they may harbor variation in the flanking regions of the amplicon [3,4]. The same concept for single nucleotide polymorphisms (SNPs) also may apply [5,6]. That is, the current SNPs in MPS kits may indeed be microhaplotypes and their amplicons may contain additional information. While Novroski, et al. [3] described STR flanking region variation in major populations, such variation in Native Americans has yet to be described. Potential genetic variation in flanking regions adjacent to 59 STRs and 172 SNPs in the ForenSeq™ DNA Signature Prep Kit's primer panel was investigated in the Yavapai Native American population. The results show a moderate degree of STR and SNP flanking region genetic variation in the Yavapai population.

* Corresponding author at: Institute of Molecular Medicine University of North Texas Health Science Center 3500 Camp Bowie Blvd, CBH–250 Fort Worth, TX 76107, USA.
E-mail address: Frank.Wendt@my.unthsc.edu (F.R. Wendt).

http://dx.doi.org/10.1016/j.fsigen.2017.02.014
1872-4973/© 2017 Elsevier B.V. All rights reserved.

2. Material and methods

MPS data for 27 autosomal, 7 X-chromosomal, and 25 Y-chromosomal STRs, Amelogenin, and 94 identity-informative SNPs (iiSNPs), 56 ancestry-informative SNPs (aiSNPs), and 22 phenotype-informative SNPs (piSNPs) were generated from DNA samples from 62 Yavapai Native Americans [7–9]. The samples and typing methods were described previously by Wendt, et al. [10]. Fastq files were used as standard input for STRait Razor v2s [3,11] to identify flanking region sequence variation.

5X depth of coverage (DoC) and 0.20 allele coverage ratio (ACR) thresholds were applied to the data. In-house Excel-based workbooks and Genetic Data Analysis (GDA) [12] were used to compile length-based and sequence-based allele frequencies, calculate observed and expected heterozygosities (H_o and H_e, respectively), perform tests for departures from Hardy-Weinberg Equilibrium (HWE) and detection of pairwise linkage disequilibrium (LD), compare previously published observed allele frequencies to the Yavapai data using the Chi-Squared Goodness of Fit Test, and generate random match probabilities (RMPs). STRUCTURE v2.3.4 (July 2012) [13–15] was used to assess the genetic structure of the Yavapai relative to the major populations described by Novroski, et al. [3] using length-based and sequence-based autosomal STRs.

A length-based STR allele is defined as only the capillary electrophoresis (CE)-based allele call and a sequence-based STR allele as the repeat motif plus flanking region sequence information, presented herein as full string sequence and condensed nomenclature consistent with the recommendations of Parson et al., 2016 [16]. SNPs do not have a length-based allele call. Here we use the term "SNP" to describe the target SNP marker only or the MPS-based marker with invariable flanking sequences and "microhaplotype" to describe an entire string sequence that includes the target SNP plus adjacent flanking region sequence variation. All SNP and microhaplotype string sequences also are paired with a condensed nomenclature consistent with recommendations by Parson et al. [16].

3. Results and discussion

The total number of loci analyzed herein includes 27 autosomal, 7 X-chromosomal, and 25 Y-chromosomal STRs, Amelogenin, and 94 iiSNPs, 56 aiSNPs, and 15 piSNPs (or piSNP-containing microhaplotypes) [3,11].

3.1. Sequencing performance

Analysis of MiSeq FGx Forensic Genomics System sequence data (with the ForenSeq™ DNA Signature Prep Kit) using STRait Razor v2s and the previously reported ForenSeq™ Universal Analysis Software (UAS) [6] produced slightly different DoCs due to the length differences of the analyzed string sequences. The average DoC and ACR using STRait Razor v2s were $1{,}600X \pm 1730$ and 0.782 ± 0.147, respectively, for 59 STRs and Amelogenin (data not shown), and $1{,}310X \pm 911$ and 0.830 ± 0.132, respectively, for 165 target SNPs and SNP microhaplotypes (data not shown).

3.2. Flanking region variation

A full discussion of length-based and repeat motif variation for ForenSeq™ DNA Signature Prep Kit STR loci in the Yavapai population can be found in Wendt, et al. [6]. It should be noted that the string sequences analyzed by STRait Razor v2s are operationally defined, based on primer placement, to maximize data recovery from the ForenSeq™ DNA Signature Prep Kit amplicon for each marker [11]. As primer sequences of commercially available MPS library preparation kits are refined or modified, the genomic coordinates of each amplicon may change, influencing the flanking region variation that can be observed for each target marker.

3.2.1. STRs

Relative to major population groups, few loci in the Yavapai population had flanking region sequence variation. Seven autosomal STRs (D13S317, D16S539, D18S51, D20S482, D2S441, D5S818, and D7S820) exhibited flanking region variation (Tables 1 and 2 and Supplemental Table 1). HWE and pairwise LD *p*-values for autosomal STRs and all other marker types, where appropriate, are listed in Supplemental Tables 2 and 3. Under the assumption of independence, the combined length-based and sequence-based RMPs for six autosomal STRs with flanking region variation were 1.80×10^{-5} and 7.80×10^{-8}, respectively.

Founder effects and population isolation of Native American populations suggest that these groups may exhibit lower genetic diversity than major population groups. Relative to Novroski, et al. [3], the Yavapai exhibited more loci with repeat region variation only and flanking region variation only than major United States populations but substantially less loci with variation in both regions (Fig. 1). Assuming the populations in Novroski, et al. [3] represent K = 4 (AFA, CHI, CAU, and HIS), STRUCTURE analysis of length-based and sequence-based autosomal STRs indicate that the Yavapai are most similar to the HIS population (data not shown).

Little flanking region variation was observed in the seven ForenSeq™ DNA Signature Prep Kit X-STR loci (Tables 1 and 2 and Supplemental Table 4). The allele, DXS10074 [CE 15]-GRCh38-ChrX:67757300-67757464 (AAGA)12 AAGG (AAGA)2 67757338-A, was observed once at the DXS100074 locus in the Yavapai. After Bonferroni correction ($p < 0.00238$), two significant pairwise LDs containing DXS10074 (DXS10074/DXS10135 and DXS10074/DXS7132) were observed ($p = 0.0015$ and 0.0021, respectively). While not observed with length-based X-STR alleles [10], significant LD between DXS10074/DXS7132 is not surprising as they are part of a linkage block [18,19].

Minimal variation was observed in the Y-STR flanking regions. Two Y-STRs (DYS389I and DYS389II) exhibited flanking region sequence variation (Tables 1 and 2 and Supplemental Table 5).

One X- and three Y-STR alleles with flanking region variation in the Yavapai population were not reported by Novroski, et al. [3] (Table 2). These novel observations may be the result of sampling error within the populations studied to date; however, given their relatively high frequencies, these alleles may be common variants within Yavapai (or other Native Americans). Future studies of larger Native American cohorts would better characterize the distribution of these alleles within and between Native American groups.

3.2.2. SNPs

To our knowledge, string sequences for target SNPs and microhaplotypes captured by the ForenSeq™ DNA Signature Prep Kit have not been published previously (Supplemental Table 6). Twenty two target iiSNPs had flanking region variation adjacent to the target SNP (Table 1). Eight (8/22) provided no increase in the number of alleles at their respective loci. For example, an iiSNP microhaplotype contains the target SNP rs1015250 (G/C) and a flanking SNP rs6475200 (A/G) (Supplemental Table 6). In this population, the target SNP G allele is always observed with the flanking SNP A allele and the target SNP C allele is always observed with the flanking SNP G allele. Four iiSNP microhaplotypes, containing the target SNPs: rs10776839, rs1109037, rs2830795, and rs876724, substantially increased in diversity compared to the target iiSNPs alone. These markers exhibited an average H_o increase of 0.318 ± 0.190, with a range of 0.0645–0.500. Assuming

Table 1
Average allele or haplotype frequencies for autosomal, X-chromosomal, and Y-chromosomal short tandem repeat (STR) and identity-informative (iiSNP), ancestry-informative (aiSNP), and phenotype-informative (piSNP) single nucleotide polymorphism (SNP) loci with flanking region sequence variation and their resulting impact on observed (H_o) and expected heterozygosities (H_e), single-locus random match probabilities (RMPs), and haplotype diversity relative to typing length-based STRs or target SNPs only. A comparison of sequence based allele frequencies from Novroski, et al. [3] was performed; *p*-values for significantly different allele frequencies are <0.05.

Marker Type	Number of Loci with Flanking Region Variation	Average Frequency	Average H_o Increase	Average H_e Increase	Average Single Locus RMP Decrease	Haplotype Frequency Decrease	Haplotype Diversity Increase	Comparison to Novroski, et al. [3] populations data	Additional Comments
Autosomal STR	7	0.0704 ± 0.0967	0.0553 ± 0.0500	0.0516 ± 0.0483	0.0400 ± 0.0409	–	–	• 26 significantly differed from all four populations • D12S391 differed from AFA, CHI, and CAU, but not HIS	• Greatest RMP decrease for D2S441 (0.13)
X-STR	1	0.0139	0	0.03	–	–	–	• DXS10103, DXS10135, DXS7423, HPRTB differed from all four populations • DXS7132 similar to all four populations • DXS10074 similar to HIS • DXS8378 similar to CHI and HIS	One novel allele observed (Table 2)
Y-STR	2	0.136 ± 0.0273	–	–	–	0	0	• DYF387S1, DYS19, DYS385ab, DYS437, DYS438, DYS481, DYS570, and DYS635 differed from AFA, CHI, CAU, and HIS • DYS389II, DYS505, DYS533, and DYS549 differed from three populations • DYS391, DYS392, DYS448, DYS576, DYS643, and Y-GATA-H4 differed from two populations • DYS389I, DYS390, DYS460, and DYS522 differed from one population • DYS439 was similar to all four	• Three novel alleles observed at DYS389I and DYS389II in the same three individuals (Table 2)
iiSNP	22	0.121 ± 0.150*	0.109 ± 0.166*	0.0865 ± 0.123*	0.0900 ± 0.115*	–	–	–	• 14/22 provided increased heterozygosity and allele spread
aiSNP	6	0.191 ± 0.215*	0.116 ± 0.144*	0.130 ± 0.171*	–	–	–	–	• 5/6 provided increased heterozygosity and allele spread, two are previously reported [17,18]
piSNP	2	0.0524 ± 0.0627	0.161	0.0895 ± 0.122	–	–	–	–	–

* Reported values are based on the indicated subset of loci showing increase in allele spread.

independence, the combined 14-locus RMPs for iiSNPs and their corresponding iiSNP microhaplotypes were 1.06×10^{-5} and 1.67×10^{-7}, respectively. The flanking region variation of 14 iiSNP microhaplotypes supports that a relatively small panel of microhaplotypes might be as informative as some low performing STR markers. Many iiSNPs lacked flanking region variation in the

Table 2
Allele frequencies, observed (H_o) and expected (H_e) heterozygosities, and Hardy-Weinberg Equilibrium (HWE) *p*-values for length-based (LB) short tandem repeats (STRs), sequence-based (SB) STRs, single nucleotide polymorphisms (SNPs), and SNP-containing microhaplotypes which exhibited an increase in number of alleles due to sequence variation in the flanking regions adjacent to the target motif. Bolded and underlined regions indicate target repeat motif and flanking sequence variation, respectively; bolded and italicized HWE *p*-values are significant after Bonferroni correction for that specific marker type.

LB Allele	Count	Frequency	String Sequence	SB Nomenclature	Count	Frequency
D13S317 (n=124)						
8	7	0.0565	TCTGACCCATCTAACGCCTATCTGTATTTACAAATACAT**TATCTATCTATCTATCTATCTATCTATCTATC**AATCAATCATCTATCTATCTTTCTGTCTGTCTTTTTGGG	D13S317 [CE 8]-GRCh38-Chr13-82147986-82148107 (TATC)8	7	0.0565
9	42	0.339	TCTGACCCATCTAACGCCTATCTGTATTTACAAATACAT**TATCTATCTATCTATCTATCTATCTATCTATCTATC**AATCAATCATCTATCTATCTTTCTGTCTGTCTTTTTGGG	D13S317 [CE 9]-GRCh38-Chr13-82147986-82148107 (TATC)9	42	0.339
10	4	0.0323	TCTGACCCATCTAACGCCTATCTGTATTTACAAATACAT**TATCTATCTATCTATCTATCTATCTATCTATCTATCTATC**TATCAATCATCTATCTATCTTTCTGTCTGTCTTTTTGGG	D13S317 [CE 10]-GRCh38-Chr13-82147986-82148107 (TATC)10 82148069-T	7	0.323
11	39	0.315	TCTGACCCATCTAACGCCTATCTGTATTTACAAATACAT**TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC**TATCAATCATCTATCTATCTTTCTGTCTGTCTTTTTGGG	D13S317 [CE 11]-GRCh38-Chr13-82147986-82148107 (TATC)11 82148069-T	11	0.0887
			TCTGACCCATCTAACGCCTATCTGTATTTACAAATACAT**TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC**AATCAATCATCTATCTATCTTTCTGTCTGTCTTTTTGGG	D13S317 [CE 11]-GRCh38-Chr13-82147986-82148107 (TATC)11	28	0.226
12	22	0.177	TCTGACCCATCTAACGCCTATCTGTATTTACAAATACAT**TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC**TATCAATCATCTATCTATCTATCTTTCTGTCTGTCTTTTTGGG	D13S317 [CE 12]-GRCh38-Chr13-82147986-82148107 (TATC)12 82148069-T	11	0.0887
			TCTGACCCATCTAACGCCTATCTGTATTTACAAATACAT**TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC**AATCAATCATCTATCTATCTATCTTTCTGTCTGTCTTTTTGGG	D13S317 [CE 12]-GRCh38-Chr13-82147986-82148107 (TATC)12	11	0.0887
13	9	0.0726	TCTGACCCATCTAACGCCTATCTGTATTTACAAATACAT**TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC**AATCAATCATCTATCTATCTTTCTGTCTGTCTTTTTGGG	D13S317 [CE 13]-GRCh38-Chr13-82147986-82148107 (TATC)13	9	0.0726
14	1	0.00806	TCTGACCCATCTAACGCCTATCTGTATTTACAAATACAT**TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC**AATCAATCATCTATCTATCTTTCTGTCTGTCTTTTTGGG	D13S317 [CE 14]-GRCh38-Chr13-82147986-82148107 (TATC)14	1	0.00806
LB H_e 0.751 LB H_o 0.710 LB HWE p-value: 0.221			SB H_e 0.808 SB H_o 0.742 SB HWE p-value: 0.0253			
D16S539 (n=124)						
9	15	0.121	TCCTCTTCCCTAGATCAATACAGACAGACAGACAGGTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAT**CATTGAAAGACAAACCAGAGATGGATGATAGATAC	D16S539 [CE 9]-GRCh38-Chr16-86352664-86352781 (GATA)9 86352761-C	12	0.0968
			TCCTCTTCCCTAGATCAATACAGACAGACAGACAGGTG**GATAGATAGATAGATAGATAGATAGATAGATAGATA**TCATTGAAAGACAAAACAGAGATGGATGATAGATAC	D16S539 [CE 9]-GRCh38-Chr16-86352664-86352781 (GATA)9	3	0.0242
10	26	0.210	TCCTCTTCCCTAGATCAATACAGACAGACAGACAGGTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**TCATTGAAAGACAAACCAGAGATGGATGATAGATAC	D16S539 [CE 10]-GRCh38-Chr16-86352664-86352781 (GATA)10 86352761-C	20	0.161
			TCCTCTTCCCTAGATCAATACAGACAGACAGACAGGTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**TCATTGAAAGACAAAACAGAGATGGATGATAGATAC	D16S539 [CE 10]-GRCh38-Chr16-86352664-86352781 (GATA)10	6	0.0484
11	30	0.242	TCCTCTTCCCTAGATCAATACAGACAGACAGACAGGTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**TCATTGAAAGACAAACCAGAGATGGATGATAGATAC	D16S539 [CE 11]-GRCh38-Chr16-86352664-86352781 (GATA)11 86352761-C	9	0.0726
			TCCTCTTCCCTAGATCAATACAGACAGACAGACAGGTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**TCATTGAAAGACAAAACAGAGATGGATGATAGATAC	D16S539 [CE 11]-GRCh38-Chr16-86352664-86352781 (GATA)11	21	0.169
12	46	0.371	TCCTCTTCCCTAGATCAATACAGACAGACAGACAGGTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**TCATTGAAAGACAAAACAGAGATGGATGATAGATAC	D16S539 [CE 12]-GRCh38-Chr16-86352664-86352781 (GATA)12	46	0.371
13	5	0.0403	TCCTCTTCCCTAGATCAATACAGACAGACAGACAGGTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**TCATTGAAAGACAAAACAGAGATGGATGATAGATAC	D16S539 [CE 13]-GRCh38-Chr16-86352664-86352781 (GATA)13	5	0.0403
14	2	0.0161	TCCTCTTCCCTAGATCAATACAGACAGACAGACAGGTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**TCATTGAAAGACAAAACAGAGATGGATGATAGATAC	D16S539 [CE 14]-GRCh38-Chr16-86352664-86352781 (GATA)14	2	0.0161
LB H_e 0.749 LB H_o 0.790 LB HWE p-value: 0.354			SB H_e 0.795 SB H_o 0.790 SB HWE p-value: 0.347			
D18S51 (n=124)						
11	1	0.00806	GTCTC**AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA**AAGAGAGAGGAAAGAAAGAGAAAAAGAAAAGAAATAGTAGCAACTGTTATTGTAAGA	D18S51 [CE 11]-GRCh38-Chr18-63281662-63281796 (AGAA)11	1	0.00806
12	10	0.0806	GTCTC**AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA**AAGAGAGAGGAAAGAAAGAGAAAAAGAAAAGAAATAGTAGCAACTGTTATTGTAAGA	D18S51 [CE 12]-GRCh38-Chr18-63281662-63281796 (AGAA)12	10	0.0806
13	22	0.177	GTCTC**AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA**AAGAGAGAGGAAAGAAAGAGAAAAAGAAAAGAAATAGTAGCAACTGTTATTGTAAGA	D18S51 [CE 13]-GRCh38-Chr18-63281662-63281796 (AGAA)13	22	0.177
14	12	0.0968	GTCTC**AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA**AAGAGAGAGGAAAGAAAGAGAAAAAGAAAAGAAATAGTAGCAACTGTTATTGTAAGA	D18S51 [CE 14]-GRCh38-Chr18-63281662-63281796 (AGAA)14	12	0.0968
15	9	0.0726	GTCTC**AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA**AAGAGAGAGGAAAGAAAGAGAAAAAGAAAAGAAATAGTAGCAACTGTTATTGTAAGA	D18S51 [CE 15]-GRCh38-Chr18-63281662-63281796 (AGAA)15	9	0.0726
16	24	0.193	GTCTC**AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA**AAGAGAGAGGAAAGAAAGAGAAAAAGAAAAGAAATAGTAGCAACTGTTATTGTAAGA	D18S51 [CE 16]-GRCh38-Chr18-63281662-63281796 (AGAA)16	24	0.194
17	31	0.250	GTCTC**AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA**AAGAGAGAGGAAAGAAAGAGAAAAAGAAAAGAAATAGTAGCAACTGTTATTGTAAGA	D18S51 [CE 17]-GRCh38-Chr18-63281662-63281796 (AGAA)17	30	0.242
			GTCTC**AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA**AGAGAGAGGAAAGAAAGAGAAAAAGAAAAGAAATAGTAGCAACTGTTATTGTAAGA	D18S51 [CE 17]-GRCh38-Chr18-63281662-63281796 (AGAA)17 63281740-G	1	0.00806
18	2	0.0161	GTCTC**AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA**AAGAGAGAGGAAAGAAAGAGAAAAAGAAAAGAAATAGTAGCAACTGTTATTGTAAGA	D18S51 [CE 18]-GRCh38-Chr18-63281662-63281796 (AGAA)18	2	0.0161
19	2	0.0161	GTCTC**AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA**AAGAGAGAGGAAAGAAAGAGAAAAAGAAAAGAAATAGTAGCAACTGTTATTGTAAGA	D18S51 [CE 19]-GRCh38-Chr18-63281662-63281796 (AGAA)19	2	0.0161
20	5	0.0403	GTCTC**AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA**AAGAGAGAGGAAAGAAAGAGAAAAAGAAAAGAAATAGTAGCAACTGTTATTGTAAGA	D18S51 [CE 20]-GRCh38-Chr18-63281662-63281796 (AGAA)20	5	0.0403
22	6	0.0484	GTCTC**AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA**AAGAGAGAGGAAAGAAAGAGAAAAAGAAAAGAAATAGTAGCAACTGTTATTGTAAGA	D18S51 [CE 22]-GRCh38-Chr18-63281662-63281796 (AGAA)22	6	0.0484
LB H_e 0.850 LB H_o 0.855 LB HWE p-value: 0.00870			SB H_e 0.854 SB H_o 0.871 SB HWE p-value: 0.0167			
D20S482 (n=124)						
12	6	0.0484	AGACACCGAACCAATAAG**AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT**AGAGATTTATTATAGGAATTGATT	D20S482 [CE 12]-GRCh38-Chr20-4525674-4525771 (AGAT)12	6	0.0484
13	28	0.226	AGACACCGAACCAATAAG**AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT**AGAGATTTATTATAGGAATTGATT	D20S482 [CE 13]-GRCh38-Chr20-4525674-4525771 (AGAT)13	27	0.218
			AGACACTGAACCAATAAG**AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT**AGAGATTTATTATAGGAATTGATT	D20S482 [CE 13]-GRCh38-Chr20-4525674-4525771 (AGAT)13 4525680-T	1	0.00807
14	73	0.589	AGACACCGAACCAATAAG**AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT**AGAGATTTATTATAGGAATTGATT	D20S482 [CE 14]-GRCh38-Chr20-4525674-4525771 (AGAT)14	67	0.540
			AGACACTGAACCAATAAG**AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT**AGAGATTTATTATAGGAATTGATT	D20S482 [CE 14]-GRCh38-Chr20-4525674-4525771 (AGAT)14 4525680-T	6	0.0484
15	14	0.113	AGACACCGAACCAATAAG**AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA**TAGAGATTTATTATAGGAATTGATT	D20S482 [CE 15]-GRCh38-Chr20-4525674-4525771 (AGAT)15	11	0.0887
			AGACACTGAACCAATAAG**AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA**TAGAGATTTATTATAGGAATTGATT	D20S482 [CE 15]-GRCh38-Chr20-4525674-4525771 (AGAT)15 4525680-T	3	0.0241
16	3	0.0242	AGACACTGAACCAATAAG**AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT**AGAGATTTATTATAGGAATTGATT	D20S482 [CE 16]-GRCh38-Chr20-4525674-4525771 (AGAT)16 4525680-T	3	0.0242
LB H_e 0.592 LB H_o 0.645 LB HWE p-value: 0.535			SB H_e 0.652 SB H_o 0.742 SB HWE p-value: 0.0744			
D2S441 (n=124)						
10	73	0.589	CCAGAAACTGTGGCTCATCTATGAAAACT**TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA**TATCATAACACCACAGCCACTTA	D2S441 [CE 10]-GRCh38-Chr2-68011918-68012017 (TCTA)10 68011922-A	20	0.161
			CCAGGAACTGTGGCTCATCTATGAAAACT**TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA**TATCATAACACCACAGCCACTTA	D2S441 [CE 10]-GRCh38-Chr2-68011918-68012017 (TCTA)10	1	0.00806
			CCAGGAACTGTGGCTCATCTATGAAAACT**TCTATCTATCTATCTATCTATCTATCTATCTATCTGTCTA**TATCATAACACCACAGCCACTTA	D2S441 [CE 10]-GRCh38-Chr2-68011918-68012017 (TCTA)8 TCTG TCTA	52	0.419
11	36	0.290	CCAGGAACTGTGGCTCATCTATGAAAACT**TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA**TATCATAACACCACAGCCACTTA	D2S441 [CE 11]-GRCh38-Chr2-68011918-68012017 (TCTA)11	35	0.282
			CCAGGAACTGTGGCTCATCTATGAAAACT**TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTGTCTA**TATCATAACACCACAGCCACTTA	D2S441 [CE 11]-GRCh38-Chr2-68011918-68012017 (TCTA)9 TCTG TCTA	1	0.00806

12	4	0.0323	CCAGGAACTGTGGCTCATCTATGAAAACT**TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA**TATCATAACACCACAGC CACTTA	D2S441 [CE 12]-GRCh38-Chr2-68011918-68012017 (TCTA)12	4	0.0323
13	1	0.00806	CCAGGAACTGTGGCTCATCTATGAAAACT**TCTATCTATCTATCTATCTATCTATCTATCTATCTATTTATCTATCTA**TATCATAACACCA CAGCCACTTA	D2S441 [CE 13]-GRCh38-Chr2-68011918-68012017 (TCTA)10 TTTA (TCTA)2	1	0.00806
14	10	0.0806	CCAGGAACTGTGGCTCATCTATGAAAACT**TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATTTATCTATCTA**TATCATAACA CCACAGCCACTTA	D2S441 [CE 14]-GRCh38-Chr2-68011918-68012017 (TCTA)11 TTTA (TCTA)2	10	0.0806
LB H_e 0.566 LB H_o 0.597 LB HWE p-value: 0.686			SB H_e 0.716 SB H_o 0.742 SB HWE p-value: 0.472			

DS5818 (n=124)						
LB Allele	Count	Frequency	String Sequence	SB Nomenclature	Count	Frequency
7	19	0.153	TATTTATACCTCT**ATCTATCTATCTATCTATCTATCTATCT**TCAAAAT	D5S818 [CE 7]-GRCh38-Chr5-123775543-123775606 (ATCT)7	19	0.153
8	7	0.0565	TATTTATACCTCT**ATCTATCTATCTATCTATCTATCTATCT**TTCAAAAT	D5S818 [CE 8]-GRCh38-Chr5-123775543-123775606 (ATCT)8	7	0.0565
9	2	0.0161	TATTTATACATCT**ATCTATCTATCTATCTATCTATCTATCT**TTCAAAAT	D5S818 [CE 9]-GRCh38-Chr5-123775543-123775606 (ATCT)9 123775552-A	2	0.0161
10	7	0.0565	TATTTATACCTCT**ATCTATCTATCTATCTATCTATCTATCTATCTATCT**TCAAAAT	D5S818 [CE 10]-GRCh38-Chr5-123775543-123775606 (ATCT)10	6	0.0484
			TATTTATACATCT**ATCTATCTATCTATCTATCTATCTATCTATCTATCT**TCAAAAT	D5S818 [CE 10]-GRCh38-Chr5-123775543-123775606 (ATCT)10 123775552-A	1	0.00806
11	75	0.605	TATTTATACCTCT**ATCTATCTATCTATCTATCTATCTATCTATCTATCT**TTCAAAAT	D5S818 [CE 11]-GRCh38-Chr5-123775543-123775606 (ATCT)11	74	0.597
			TATTTATACATCT**ATCTATCTATCTATCTATCTATCTATCTATCTATCT**TTCAAAAT	D5S818 [CE 11]-GRCh38-Chr5-123775543-123775606 (ATCT)11 123775552-A	1	0.00806
12	13	0.105	TATTTATACCTCT**ATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT**TTCAAAAT	D5S818 [CE 12]-GRCh38-Chr5-123775543-123775606 (ATCT)12	10	0.0806
			TATTTATACATCT**ATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT**TTCAAAAT	D5S818 [CE 12]-GRCh38-Chr5-123775543-123775606 (ATCT)12 123775552-A	3	0.0242
13	1	0.00806	TATTTATACATCT**ATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT**TCAAAAT	D5S818 [CE 13]-GRCh38-Chr5-123775543-123775606 (ATCT)13 123775552-A	1	0.00806
LB H_e 0.598 LB H_o 0.532 LB HWE p-value: 0.263			SB H_e 0.612 SB H_o 0.581 SB HWE p-value: 0.445			

D7S820 (n=124)						
LB Allele	Count	Frequency	String Sequence	SB Nomenclature	Count	Frequency
7	1	0.00806	TATTTAGTGAGATAAAAAAAACTATCAATCTGTC**TATCTATCTATCTATCTATCTATCTATC**GTTAGTTCATTCTAAACTAT	D7S820 [CE 7]-GRCh38-Chr7-84160191-84160297 (TATC)7 84160204-A; 84160286-A	1	0.00806
8	7	0.0565	TATTTAGTGAGATAAAAAAAACTATCAATCTGTC**TATCTATCTATCTATCTATCTATCTATC**GTTAGTTCATTCTAAACTAT	D7S820 [CE 8]-GRCh38-Chr7-84160191-84160297 (TATC)8 84160204-A; 84160286-A	1	0.00806
			TATTTAGTGAGATAAAAAAAACTATCAATCTGTC**TATCTATCTATCTATCTATCTATCTATC**GTTAGTTCGTTCTAAACTAT	D7S820 [CE 8]-GRCh38-Chr7-84160191-84160297 (TATC)8 84160204-A	6	0.0484
9	5	0.0403	TATTTAGTGAGATAAAAAAAACTATCAATCTGTC**TATCTATCTATCTATCTATCTATCTATC**GTTAGTTCGTTCTAAACTAT	D7S820 [CE 9]-GRCh38-Chr7-84160191-84160297 (TATC)9 84160204-A	5	0.0403
10	15	0.121	TATTTAGTGAGATAAAAAAAACTATCAATCTGTC**TATCTATCTATCTATCTATCTATCTATC**GTTAGTTCGTTCTAAACTA T	D7S820 [CE 10]-GRCh38-Chr7-84160191-84160297 (TATC)10 84160204-A	15	0.121
11	53	0.427	TATTTAGTGAGATAAAAAAAACTATCAATCTGTC**TATCTATCTATCTATCTATCTATCTATC**GTTAGTTCGTTCTAA ACTAT	D7S820 [CE 11]-GRCh38-Chr7-84160191-84160297 (TATC)11 84160204-A	52	0.419
			TATTTAGTGAGATTAAAAAAAACTATCAATCTGTC**TATCTATCTATCTATCTATCTATCTATC**GTTAGTTCGTTCTAAA CTAT	D7S820 [CE 11]-GRCh38-Chr7-84160191-84160297 (TATC)11	1	0.00806
12	40	0.323	TATTTAGTGAGATAAAAAAAACTATCAATCTGTC**TATCTATCTATCTATCTATCTATCTATC**GTTAGTTCGTTC TAAACTAT	D7S820 [CE 12]-GRCh38-Chr7-84160191-84160297 (TATC)12 84160204-A	35	0.283
			TATTTAGTGAGATTAAAAAAAACTATCAATCTGTC**TATCTATCTATCTATCTATCTATCTATC**GTTAGTTCGTTC TAAACTAT	D7S820 [CE 12]-GRCh38-Chr7-84160191-84160297 (TATC)12	5	0.0403
13	1	0.00806	TATTTAGTGAGATAAAAAAAACTATCAATCTGTC**TATCTATCTATCTATCTATCTATCTATC**GTTAGTTC GTTCTAAACTAT	D7S820 [CE 13]-GRCh38-Chr7-84160191-84160297 (TATC)13 84160204-A	1	0.00806
14	2	0.0161	TATTTAGTGAGATAAAAAAAACTATCAATCTGTC**TATCTATCTATCTATCTATCTATCTATC**GTTAGTTCGTTA GTTCGTTCTAAACTAT	D7S820 [CE 14]-GRCh38-Chr7-84160191-84160297 (TATC)14 84160204-A	2	0.0161
LB H_e 0.699 LB H_o 0.726 LB HWE p-value: 0.456			SB H_e 0.730 SB H_o 0.774 SB HWE p-value: 0.613			

DXS10074 (n♂=72,n♀=26)						
LB Allele	Count (f|m)	Frequency (f|m)	String Sequence	SB Nomenclature	Count (f|m)	Frequency (f|m)
14	5|1	0.0694|0.0385	TGTGTGTGCATGCATACACACACAGAGAGAGAGAGAGAGAAAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA GAAAGGAAGAAAGAAAGGAAGAAAAATAGAACAAATCAGCTTATATTCAGTATTTTTTTAGTATTTTCTGTGTCAGCTC	DXS10074 [CE 14]-GRCh38-ChrX-67757300-67757464 (AAGA)11 AAGG (AAGA)2	5|1	0.06940|0.0385
15	20|4	0.278|0.154	TGTGTGTGCATGCATACACACACAGAGAGAGAGAGAGAGAAAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA GAAAGAAAGGAAGAAAGAAAGGAAGAAAAATAGAACAAATCAGCTTATATTCAGTATTTTTTTAGTATTTTCTGTGTCAGCTC	DXS10074 [CE 15]-GRCh38-ChrX-67757300-67757464 (AAGA)12 AAGG (AAGA)2	16|3	0.222|0.115
			TGTGTGTGCATGCATACACACACAGAGAGAGAGAGAGAGAAAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA GAAAGAAAGGAAGAAAGAAAGGAAGAAAAATAGAACAAATCAGCTTATATTCAGTATTTTTTTAGTATTTTCTGTGTCAGCTC	DXS10074 [CE 15]-GRCh38-ChrX-67757300-67757464 (AAGA)15	3|1	0.0417|0.0385
			TGTGTGTGCATGCATACACACACAGAGAGAGAGAGAGAGAAAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA GAAAGAAAGGAAGAAAGAAAGGAAGAAAAATAGAACAAATCAGCTTATATTCAGTATTTTTTTAGTATTTTCTGTGTCAGCTC	DXS10074 [CE 15]-GRCh38-ChrX-67757300-67757464 (AAGA)12 AAGG (AAGA)2 67757338-A	1|0	0.0139|0
16	13|9	0.181|0.346	TGTGTGTGCATGCATACACACACAGAGAGAGAGAGAGAGAAAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA GAAAGAAAGAAAGGAGGAAAGAAAGGAAGAAAATAGAACAAATCAGCTTATATTCAGTATTTTTTTAGTATTTTCTGTGTCAGCTC	DXS10074 [CE 16]-GRCh38-ChrX-67757300-67757464 (AAGA)13 AAGG AGGA AAGA	0|2	0|0.0769
			TGTGTGTGCATGCATACACACACAGAGAGAGAGAGAGAGAAAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA GAAAGAAAGAAAGGAAGAAAGAAAGGAAGAAAATAGAACAAATCAGCTTATATTCAGTATTTTTTTAGTATTTTCTGTGTCAGCTC	DXS10074 [CE 16]-GRCh38-ChrX-67757300-67757464 (AAGA)13 AAGG (AAGA)2	13|7	0.181|0.269
17	22|7	0.306|0.269	TGTGTGTGCATGCATACACACACAGAGAGAGAGAGAGAGAAAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA GAAAGAAAGAAAGAAAGGAAGAAAGAAAGGAAGAAAATAGAACAAATCAGCTTATATTCAGTATTTTTTTAGTATTTTCTGTGTCAGCTC	DXS10074 [CE 17]-GRCh38-ChrX-67757300-67757464 (AAGA)14 AAGG (AAGA)2	22|7	0.306|0.269
18	9|3	0.125|0.115	TGTGTGTGCATGCATACACACACAGAGAGAGAGAGAGAGAAAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA GAAAGAAAGAAAGAAAGGAAGAAAGAAAGGAAGAAAATAGAACAAATCAGCTTATATTCAGTATTTTTTTAGTATTTTCTGTGTCAG CTC	DXS10074 [CE 18]-GRCh38-ChrX-67757300-67757464 (AAGA)15 AAGG (AAGA)2	9|3	0.125|0.115
19	3|2	0.0417|0.0769	TGTGTGTGCATGCATACACACACAGAGAGAGAGAGAGAGAAAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA GAAAGAAAGAAAGAAAGAAAGAAAGGAAGAAAGAAAGGAAGAAAATAGAACAAATCAGCTTATATTCAGTATTTTTTTAGTATTTTCTGT GTCAGCTC	DXS10074 [CE 19]-GRCh38-ChrX-67757300-67757464 (AAGA)16 AAGG (AAGA)2	3|2	0.0417|0.0769
LB H_e* 0.786 LB H_o* 0.806 LB HWE p-value: 0.337			SB H_e* 0.812 SB H_o* 0.806 SB HWE p-value: 0.486			
*Does not include male X-STR data due to hemizygosity of the male sex chromosome.						

DYS389I (n=26)						
LB Allele	Count	Frequency	String Sequence	SB Nomenclature	Count	Frequency
12	6	0.231	**TAGATAGATAGATAGATAGATAGATAGATAGATAGACAGACAGACAGA**TACATAGATAATACAGAT	DYS389I [CE 12]-GRCh38-ChrY-12500448-12500513 (TAGA)9 (CAGA)3	6	0.231
13	18	0.692	**TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGACAGACAGACAGA**TACATAGATAATACAGAT	DYS389I [CE 13]-GRCh38-ChrY-12500448-12500513 (TAGA)10 (CAGA)3	15	0.577
			TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGACAGACAGACAGATACATAGATAGTACAGAT	DYS389I [CE 13]-GRCh38-ChrY:12500448-12500513 (TAGA)10 (CAGA)3 12500506-G	3	0.115
14	2	0.0769	**TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGACAGACAGACAGA**TACATAGATAATACAGAT	DYS389I [CE 14]-GRCh38-ChrY-12500448-12500513 (TAGA)11 (CAGA)3	2	0.0769

DYS389II* (n=24)						
LB Allele	Count	Frequency	String Sequence	SB Nomenclature	Count	Frequency
28	1	0.0417	**TAGATAGATAGATAGATAGATAGATAGATAGATAGACAGACAGACAGA**TACATAGATAATACAGATGAGAGTTGGATACAGAAGTAGG TATAATG**ATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA**TACACACATAGATAAT ACAGAT	DYS389II [CE 28]-GRCh38-ChrY:12500448-12500533 (TAGA)9 (CAGA)3 TACA TAGA TAAT ACAG ATGA GAGT TGGA TACA GAAG TAGG TATA ATGA (TAGA)11 (CAGA)5	1	0.0417
29	9	0.375	**TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGACAGACAGACAGA**TACATAGATAATACAGATGAGAGTTGGATACAGAAGTAGG TATAATG**ATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA**TACACACATAGA TAATACAGAT	DYS389II [CE 29]-GRCh38-ChrY:12500448-12500533 (TAGA)9 (CAGA)3 TACA TAGA TAAT ACAG ATGA GAGT TGGA TACA GAAG TAGG TATA ATGA (TAGA)12 (CAGA)5	3	0.125
			TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGACAGACAGACAGATACATAGATAATACAGATGAGAGTTGGATACAGAAG TAGGTATAATG**ATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA**TACACACATAGA TAATACAGAT	DYS389II [CE 29]-GRCh38-ChrY:12500448-12500533 (TAGA)10 (CAGA)3 TACA TAGA TAAT ACAG ATGA GAGT TGGA TACA GAAG TAGG TATA ATGA (TAGA)11 (CAGA)5	3	0.125
			TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGACAGACAGACAGATACATAGATAATACAGATGAGAGTTGGATACAGAAG TAGGTATAATG**ATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA**TACACACATAGA TAATACAGAT	DYS389II [CE 29]-GRCh38-ChrY:12500448-12500533 (TAGA)10 (CAGA)3 TACA TAGA TAAT ACAG ATGA GAGT TGGA TACA GAAG TAGG TATA ATGA (TAGA)12 (CAGA)4	2	0.0833
30	12	0.500	**TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGACAGACAGACAGA**TACATAGATAATACAGATGAGAGTTGGATACAGAAG TAGGTATAATG**ATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA**TACACACATA GATAATACAGAT	DYS389II [CE 30]-GRCh38-ChrY:12500448-12500533 (TAGA)10 (CAGA)3 TACA TAGA TAAT ACAG ATGA GAGT TGGA TACA GAAG TAGG TATA ATGA (TAGA)12 (CAGA)5	7	0.292
			TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGACAGACAGACAGATACATAGATAGTACAGATGAGAGTTGGATACAGAAG TAGGTATAATG**ATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA**TACACACA TAGATAATACAGAT	DYS389II [CE 30]-GRCh38-ChrY:12500448-12500533 (TAGA)10 (CAGA)3 TACA TAGA TAGT ACAG ATGA GAGT TGGA TACA GAAG TAGG TATA ATGA (TAGA)12 (CAGA)5 12500506-G	3	0.125

31	2	0.0833	**TAGATAGATAGATAGATAGATAGATAGATAGATAGACAGACAGACAGA**TACATAGATAATACAGATGAGAGTTGGATACAGAAG TAGGTATAATGA**TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGACAGACAGACAGACAGACAGA**CAC ACACATAGATAATACAGAT **TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGACAGACAGACAGA**TACATAGATAATACAGATGAGAGTTGGATACA GAAGTAGGTATAATGAT**AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGACAGACAGACAGACAGACAGA**CAC ACACATAGATAATACAGAT	DYS389II [CE 31]-GRCh38-ChrY:12500448-12500633 (TAGA)10 (CAGA)3 TACA TAGA TAAT ACAG ATGA GAGT TGGA TACA GAAG TAGG TATA ATGA (TAGA)13 (CAGA)5 DYS389II [CE 31]-GRCh38-ChrY:12500448-12500633 (TAGA)11 (CAGA)3 TACA TAGA TAAT ACAG ATGA GAGT TGGA TACA GAAG TAGG TATA ATGA (TAGA)12 (CAGA)5	1 1	0.0417 0.0417

*Flanking region variation observed within the DYS389II locus is the same as that observed within the DYS389I locus due to use of the same forward primer.

rs10776839 (n=124)					
Target Allele	Count	Frequency	String Sequence	Microhaplotype Nomenclature	Count	Frequency
T	61	0.492	CAGCTGAGGAGCCC**G**AG**T**TGCCGTCAGATCAGAGCCCCAGTTGCCCGGTCTGCCGCAGCTCT	rs10776839 [CE T]-GRCh38-chr9:134525445-134525507 rs7037930-G; rs10776839-T	61	0.492
G	63	0.508	CAGCTGAGGAGCCC**G**AG**G**TTGCCGTCAGATCAGAGCCCCAGTTGCCCGGTCTGCCGCAGCTCT	rs10776839 [CE G]-GRCh38-chr9:134525445-134525507 rs7037930-G; rs10776839-G	42	0.339
			CAGCTGAGGAGCCC**A**AG**G**TTGCCGTCAGATCAGAGCCCCAGTTGCCCGGTCTGCCGCAGCTCT	rs10776839 [CE G]-GRCh38-chr9:134525445-134525507 rs10776839-G	21	0.169
Target SNP H_e 0.504			Microhaplotype H_e 0.620			
Target SNP H_o 0.468			Microhaplotype H_o 0.532			
Target SNP HWE p-value: 0.619			Microhaplotype HWE p-value: 0.172			

rs1109037 (n=124)					
Target Allele	Count	Frequency	String Sequence	Microhaplotype Nomenclature	Count	Frequency
A	29	0.234	CCAGTTTCTCC**A**GAGTGGAAAGACTTTCATCTCGCACTGGCACGACCTTGAGACCCCGGGTTCTGATGAACTGGG**A**GG	rs1109037 [CE A]-GRCh38-chr2:9945582-9945659 rs1109037-A; rs1109038-A	19	0.153
			CCAGTTTCTCC**A**GAGTGGAAAGACTTTCATCTCGCACTGGCACGACCTTGAGACCCCGGGTTCTGATGAACTGGG**G**GG	rs1109037 [CE A]-GRCh38-chr2:9945582-9945659 rs1109037-A	10	0.0806
G	95	0.766	CCAGTTTCTCC**G**GAGTGGAAAGACTTTCATCTCGCACTGGCACGACCTTGAGACCCCGGGTTCTGATGAACTGGG**A**GG	rs1109037 [CE G]-GRCh38-chr2:9945582-9945659 rs1109037-G; rs1109038-A	49	0.395
			CCAGTTTCTCC**G**GAGTGGAAAGACTTTCATCTCGCACTGGCACGACCTTGAGACCCCGGGTTCTGATGAACTGGG**G**GG	rs1109037 [CE G]-GRCh38-chr2:9945582-9945659 rs1109037-G	46	0.371
Target SNP H_e 0.361			Microhaplotype H_e 0.682			
Target SNP H_o 0.403			Microhaplotype H_o 0.694			
Target SNP HWE p-value: 0.486			Microhaplotype HWE p-value: 0.728			

rs12997453 (n=124)					
Target Allele	Count	Frequency	String Sequence	Microhaplotype Nomenclature	Count	Frequency
G	88	0.709677	TTTTATGCTTTAAAGATA**C**AGGTTATCTGTATTACATTG**G**GTTTTTACCTACCTT	rs12997453 [CE G]-GRCh38-chr2:181548493-181548547 rs12997453-G	88	0.710
A	36	0.290323	TTTTATGCTTTAAAGATA**T**AGGTTATCTGTATTACATTG**A**GTTTTTACCTACCTT	rs12997453 [CE A]-GRCh38-chr2:181548493-181548547 rs72883670-T; rs12997453-A	27	0.218
			TTTTATGCTTTAAAGATA**C**AGGTTATCTGTATTACATTG**A**GTTTTTACCTACCTT	rs12997453 [CE A]-GRCh38-chr2:181548493-181548547 rs12997453-A	9	0.0726
Target SNP H_e 0.415			Microhaplotype H_e 0.447			
Target SNP H_o 0.387			Microhaplotype H_o 0.468			
Target SNP HWE p-value: 0.755			Microhaplotype HWE p-value: 0.0922			

rs13218440 (n=124)					
Target Allele	Count	Frequency	String Sequence	Microhaplotype Nomenclature	Count	Frequency
A	60	0.484	CCTCTG**A**GCAGCCTCCTGGAATACTCAGCTGGGATGGGTTGGGGCTGCTTGAGGTACAGCTCCCACT**G**CCTCTGAGTGGCCCTCCATGAAA ATGCCTCATGTCTCTGTGTCCCTAAACTGTAGG	rs13218440 [CE A]-GRCh38-chr6:12059715-12059838 rs13218440-A	60	0.483871
G	64	0.516	CCTCTG**G**GCAGCCTCCTGGAATACTCAGCTGGGATGGGTTGGGGCTGCTTGAGGTACAGCTCCCACT**G**CCTCTGAGTGGCCCTCCATGAAA ATGCCTCATGTCTCTGTGTCCCTAAACTGTAGG	rs13218440 [CE G]-GRCh38-chr6:12059715-12059838 rs13218440-G	63	0.508065
			CCTCTG**G**GCAGCCTCCTGGAATACTCAGCTGGGATGGGTTGGGGCTGCTTGAGGTACAGCTCCCACT**A**CCTCTGAGTGGCCCTCCATGAAA ATGCCTCATGTCTCTGTGTCCCTAAACTGTAGG	rs13218440 [CE G]-GRCh38-chr6:12059715-12059838 rs13218440-G,12059782-A	1	0.00806452
Target SNP H_e 0.504			Microhaplotype H_e 0.512			
Target SNP H_o 0.516			Microhaplotype H_o 0.532			
Target SNP HWE p-value: 1.00			Microhaplotype HWE p-value: 0.797			

rs1360288 (n=124)					
Target Allele	Count	Frequency	String Sequence	Microhaplotype Nomenclature	Count	Frequency
C	100	0.806	GGGGAGAGGCTCCCTGCAGCCTCTGGGGAGTAGAGGAGACCTGGGAAGA**C**ACTGGCTGCATCCCCTCAACAGATGCCCC	rs1360288 [CE C]-GRCh38-chr9:126205735-126205813 rs1360288-C	99	0.798
			GGG**T**AGAGGCTCCCTGCAGCCTCTGGGGAGTAGAGGAGACCTGGGAAGA**C**ACTGGCTGCATCCCCTCAACAGATGCCCC	rs1360288 [CE C]-GRCh38-chr9:126205735-126205813 rs116412791-T; rs1360288-C	1	0.00806
T	24	0.194	GGGGAGAGGCTCCCTGCAGCCTCTGGGGAGTAGAGGAGACCTGGGAAGA**T**ACTGGCTGCATCCCCTCAACAGATGCCCC	rs1360288 [CE T]-GRCh38-chr9:126205735-126205813 rs1360288-T	24	0.194
Target SNP H_e 0.315			Microhaplotype H_e 0.328			
Target SNP H_o 0.322			Microhaplotype H_o 0.339			
Target SNP HWE p-value: 1.00			Microhaplotype HWE p-value: 1.00			

rs2830795 (n=124)					
Target Allele	Count	Frequency	String Sequence	Microhaplotype Nomenclature	Count	Frequency
G	29	0.234	GAGACTGGGTTCAC**T**TCTATAGACATAGGACACACCATTTTATTGTCTAAAG**G**GCAAAGAAGTCCTATTA	rs2830795 [CE G]-GRCh38-chr21:27235792-27235861 rs2830795-G	29	0.234
A	95	0.766	GAGACTGGGTTCAC**T**TCTATAGACATAGGACACACCATTTTATTGTCTAAAG**A**GCAAAGAAGTCCTATTA	rs2830795 [CE A]-GRCh38-chr21:27235792-27235861 rs2830795-A	49	0.395
			GAGACTGGGTTCAC**C**TCTATAGACATAGGACACACCATTTTATTGTCTAAAG**A**GCAAAGAAGTCCTATTA	rs2830795 [CE A]-GRCh38-chr21:27235792-27235861 rs12626695-C; rs2830795-A	42	0.339
			GAGACTGGGTTCAC**C**TCTATAGACATAGGACACACCATTTTATTGTCTAAAG**A**TCAAAGAAGTCCTATTA	rs2830795 [CE A]-GRCh38-chr21:27235792-27235861 rs12626695-C; 27235845-T; rs2830795-A	4	0.0323
Target SNP H_e 0.361			Microhaplotype H_e 0.679			
Target SNP H_o 0.274			Microhaplotype H_o 0.694			
Target SNP HWE p-value: 0.0749			Microhaplotype HWE p-value: 0.101			

rs354439 (n=124)					
Target Allele	Count	Frequency	String Sequence	Microhaplotype Nomenclature	Count	Frequency
A	104	0.839	TTCTTAACTCTCAAATT**G**CAGGTTGCGATAGAAAACAGTGAATGATATTCAGAATATTGTGCTTAA**A**AGGAATGTGATACATGAGAGAGAT ACATAAGGGAAAGAGAAGCCACCAGAACA	rs354439 [CE A]-GRCh38-chr13:106285996-106286115 rs354439-A	103	0.831
			TTCTTAACTCTCAAATT**T**CAGGTTGCGATAGAAAACAGTGAATGATATTCAGAATATTGTGCTTAAAAGGAATGTGATACATGAGAGAGATA CATAAGGGAAAGAGAAGCCACCAGAACA	rs354439 [CE A]-GRCh38-chr13:106285996-106286115 106286013-T; rs354439-A	1	0.00806
T	20	0.161	TTCTTAACTCTCAAATT**G**CAGGTTGCGATAGAAAACAGTGAATGATATTCAGAATATTGTGCTTAATAGGAATGTGATACATGAGAGAGATA CATAAGGGAAAGAGAAGCCACCAGAACA	rs354439 [CE T]-GRCh38-chr13:106285996-106286115 rs354439-T	20	0.161
Target SNP H_e 0.273			Microhaplotype H_e 0.286			
Target SNP H_o 0.258			Microhaplotype H_o 0.274			
Target SNP HWE p-value: 0.644			Microhaplotype HWE p-value: 0.705			

rs4606077 (n=124)					
Target Allele	Count	Frequency	String Sequence	Microhaplotype Nomenclature	Count	Frequency
T	34	0.274	TGGGATCTGACTCCCCACAGCCTACCCAAAGC**T**GGGGAACTCCTCACTGCCCTTCGGGCTTCAGCACAGGGCTGTCTCCCACGCCGGCAGG GCCTGTGCTTTCACTGG	rs4606077 [CE T]-GRCh38-chr8:143574562-143574669 rs4606077-T; rs58774517-T	1	0.00806
			TGGGATCTGACTCCCCACAGCCTACCCAAAGC**C**GGGGAACTCCTCACTGCCCTTCGGGCTTCAGCACAGGGCTGTCTCCCACGCCGGCAGG GCCTGTGCTTTCACTGG	rs4606077 [CE T]-GRCh38-chr8:143574562-143574669 rs4606077-T	33	0.2661
C	90	0.726	TGGGATCTGACTCCCCACAGCCC**A**CCCAAAGC**A**GGGGAACTCCTCACTGCCCTTCGGGCTTCAGCACAGGGCTGTCTCCCACGCCGGCAGG GCCTGTGCTTTCACTGG	rs4606077 [CE C]-GRCh38-chr8:143574562-143574669 rs4606077-C; rs1869434-A	90	0.726
Target SNP H_e 0.401			Microhaplotype H_e 0.406			
Target SNP H_o 0.387			Microhaplotype H_o 0.387			
Target SNP HWE p-value: 0.764			Microhaplotype HWE p-value: 0.816			

rs719366 (n=124)					
Target Allele	Count	Frequency	String Sequence	Microhaplotype Nomenclature	Count	Frequency
A	97	0.782	GAGTAAAGGACAGGCCCCCAGAGGAGTAGCT**A**CTAGAATGGGAGGAGGAAAAGCAGGAAAGGATAATAAAAGAGTTAAAAGATGCTGT GGTCACAGTGCTGAATGCTGCCGAGAAGTGAC	rs719366 [CE A]-GRCh38-chr19:27972398-27972517 rs719366-A	97	0.782
G	27	0.218	GAGTAAAGGACAGGCCCCCAGAGGAGTAGCT**G**CTAGAATGGGAGGAGGAAAAGCAGGAAAGGATAATAAAAGAGTTAAAAGATGCTGT GGTCACAGTGCTGAATGCTGCC**A**AGAAGTGAC	rs719366 [CE G]-GRCh38-chr19:27972398-27972517 rs719366-G; rs719367-A	2	0.0161
			GAGTAAAGGACAGGCCCCCAGAGGAGTAGCT**G**CTAGAATGGGAGGAGGAAAAGCAGGAAAGGATAATAAAAGAGTTAAAAGATGCTGT GGTCACAGTGCTGAATGCTGCC**G**AGAAGTGAC	rs719366 [CE G]-GRCh38-chr19:27972398-27972517 rs719366-G	25	0.202
Target SNP H_e 0.343			Microhaplotype H_e 0.350			
Target SNP H_o 0.274			Microhaplotype H_o 0.274			
Target SNP HWE p-value: 0.1350			Microhaplotype HWE p-value: 0.155			

rs876724 (n=124)					
Target Allele	Count	Frequency	String Sequence	Microhaplotype Nomenclature	Count	Frequency
T	30	0.242	AGTATATTTTTTGTCACACTCTGCTAACTGCCTGCTCATAGATATTCAAATTTAGTAGATGTAGA**T**A	rs876724 [CE T]-GRCh38-chr2:114970-115036 rs876724-T; rs300773-T	30	0.241935

C	94	0.758	AGTACATTTTTTGTCACACTCTGCTAACTGCCTGCTCATAGATATTCAAATTTAGTAGATGTAGATA	rs876724 [CE C]-GRCh38-chr2:114970-115036 rs876724-C; rs300773-T	39	0.314516
			AGTACATTTTTTGTCACACTCTGCTAACTGCCTGCTCATAGATATTCAAATTTAGTAGATGTAGACA	rs876724 [CE C]-GRCh38-chr2:114970-115036 rs876724-C	55	0.443548
Target SNP H_e 0.370			Microhaplotype H_e 0.651			
Target SNP H_o 0.258			Microhaplotype H_o 0.758			
Target SNP HWE p-value: 0.0263			Microhaplotype HWE p-value: < ***0.000532***			

rs907100 (n=124)						
Target Allele	Count	Frequency	String Sequence	Microhaplotype Nomendature	Count	Frequency
C	40	0.323	TGATGCCCTGGCATCAAAGAAGGCTCCAACTGGGCTTCTTTTCTGTGTTTTCCAAGGCTTGGAAAG	rs907100 [CE C]-GRCh38-chr2:238654924-238654989 rs907100-C	35	0.282
			TGATGCCCTGGCATCAAAGAAGGCTCCAACTGAGCTTCTTTTCTGTGTTTTCCAAGGCTTGGAAAG	rs907100 [CE C]-GRCh38-chr2:238654924-238654989 rs907100-C; rs11689319-A	5	0.0403
G	84	0.677	TGATGCCCTGGCATGAAAGAAGGCTCCAACTGGGCTTCTTTTCTGTGTTTTCCAAGGCTTGGAAAG	rs907100 [CE G]-GRCh38-chr2:238654924-238654989 rs907100-G	84	0.677
Target SNP H_e 0.441			Microhaplotype H_e 0.464			
Target SNP H_o 0.484			Microhaplotype H_o 0.516			
Target SNP HWE p-value: 0.561			Microhaplotype HWE p-value: 0.589			

rs987640 (n=124)						
Target Allele	Count	Frequency	String Sequence	Microhaplotype Nomendature	Count	Frequency
A	51	0.411	ACAGGTACATTCACTTAACAGGCTCTCTTTCCACCCATGTAGAAATACAAAAATAAGACTTAATACAGACGATGG	rs987640 [CE A]-GRCh38-chr22:33163486-33163560 rs987640-A	50	0.403
			ACCGGTACATTCACTTAACAGGCTCTCTTTCCACCCATGTAGAAATACAAAAATAAGACTTAATACAGACGATGG	rs987640 [CE A]-GRCh38-chr22:33163486-33163560 rs17793354-C; rs987640-A	1	0.00806
T	73	0.589	ACAGGTACATTCACTTAACAGGCTCTCTTTCCACCCTTGTAGAAATACAAAAATAAGACTTAATACAGACGATGG	rs987640 [CE T]-GRCh38-chr22:33163486-33163560 rs987640-T	73	0.589
Target SNP H_e 0.488			Microhaplotype H_e 0.495			
Target SNP H_o 0.468			Microhaplotype H_o 0.484			
Target SNP HWE p-value: 0.800			Microhaplotype HWE p-value: 0.660			

rs9905977 (n=124)						
Target Allele	Count	Frequency	String Sequence	Microhaplotype Nomendature	Count	Frequency
A	48	0.387	TGGTGTCCAAGGAGGGCTGGGTGACTCGTGGCTCAGTCAGCATCAAGATTCCTTTCGTCTTTCCCCTCTGCCCTCCCTGGCTTGTCAGCTTTG TCCCTCAGGCTTGGCCCCCTCGTGGCC	rs9905977 [CE A]-GRCh38-chr17:3016058-3016176 rs9905977-A	48	0.387
G	76	0.613	TGGTGTCCAAGGAGGGCTGGGTGACTCGTGGCTCAGTCAGCGTCAAGATTCCTTTCGTCTTTCCCCTCTGCCCTCCCTAGCTTGTCAGCTTTG TCCCTCAGGCTTGGCCCCCTCGTGGCC	rs9905977 [CE G]-GRCh38-chr17:3016058-3016176 rs9905977-G; rs28582109-A	3	0.0242
			TGGTGTCCAAGGAGGGCTGGGTGACTCGTGGCTCAGTCAGCGTCAAGATTCCTTTCGTCTTTCCCCTCTGCCCTCCCTGGCTTGTCAGCTTTG TCCCTCAGGCTTGGCCCCCTCGTGGCC	rs9905977 [CE G]-GRCh38-chr17:3016058-3016176 rs9905977-G	71	0.573
			TGGTGTCCAAGGAGGGCTGGGTGACTCGTGGCTCAGTCAGCGTCAAGATTCCTTTCGTCTTTCCCCTCTGCCCTCCCTGGCTTGTCAGCTTTG TCCCTCAGGCTTGGCCCTCTCGTGGCC	rs9905977 [CE G]-GRCh38-chr17:3016058-3016176 rs9905977-G; rs73298992-T	2	0.0161
Target SNP H_e 0.478			Microhaplotype H_e 0.526			
Target SNP H_o 0.516			Microhaplotype H_o 0.565			
Target SNP HWE p-value: 0.593			Microhaplotype HWE p-value: 0.561			

rs993934 (n=124)						
Target Allele	Count	Frequency	String Sequence	Microhaplotype Nomendature	Count	Frequency
A	85	0.685	TTTGCTTTGTAAGGCAATAGAGCAAAGTATTGTGATAACAGTCTCCAGAGTATATTAGCTTAGTTCATAA	rs993934 [CE A]-GRCh38-chr2:123351571-123351640 rs993934-A	83	0.669
			TTTGCTTTGTAAGGCAATAGAGCAAAGTATTGCGATAACAGTCTCCAGAGTATATTAGCTTAGTTCATAA	rs993934 [CE A]-GRCh38-chr2:123351571-123351640 123351603-C; rs993934-A	2	0.0161
G	39	0.315	TTTGCTTTGTAAGGCAATAGAGCAAAGTATTGTGATAACAGTCTCCAGAGTATATTAGCTTAGTTCGTAA	rs993934 [CE G]-GRCh38-chr2:123351571-123351640 rs993934-G	39	0.315
Target SNP H_e 0.435			Microhaplotype H_e 0.456			
Target SNP H_o 0.565			Microhaplotype H_o 0.597			
Target SNP HWE p-value: 0.0184			Microhaplotype HWE p-value: 0.0250			

rs200354 (n=124)						
Target Allele	Count	Frequency	String Sequence	Microhaplotype Nomendature	Count	Frequency
G	12	0.0968	CAAGCTGCCTTGGAACTGGGCTGCCCCCATGCACCATGGCCATTTGGAACTGGGTAGTGAGAGGCTGCCCTGTCCATTGTAGAATGTTTAG CAGCATCCTCTGGGAGTAGCAAACTCCCCTC	rs200354 [CE G]-GRCh38-chr14:98908931-98909052 rs200354-G	12	0.0968
T	112	0.903	CAAGCTGCCTTGGAACTGGGCTGCCCCCATGCACCATGGCCATTTGGAACTGGTAGTGAGAGGCTGCCCTGTCCGTTGTAGAATGTTTAG CAGCATCCTCTGGGAGTAGCAAACTCCCCTC	rs200354 [CE T]-GRCh38-chr14:98908931-98909052 rs200354-T; rs200353-G	72	0.581
			CAAGCTGCCTTGGAACTGGGCTGCCCCCATGCACCATGGCCATTTGGAACTGGTAGTGAGAGGCTGCCCTGTCCATTGTAGAATGTTTAG CAGCATCCTCTGGGAGTAGCAAACTCCCCTC	rs200354 [CE T]-GRCh38-chr14:98908931-98909052 rs200354-T	23	0.185

Yavapai population, however, access to these sequence data may reveal additional variation on a case-by-case basis or in other population groups, defining additional microhaplotypes.

Six aiSNPs had sequence variation adjacent to the target locus with an average frequency of 0.191 ± 0.215 (Table 1 and Supplemental Table 5). Two of the aiSNP microhaplotypes (rs1079597-rs1079598 or "mh11KK-090"; rs870347-870348 or "mh11KK-062" or "PAPD7") have been described previously [17,18]. One aiSNP microhaplotype (rs1079597-rs1079598; mh11KK-090) provided no increase in the number of alleles at the target rs1079597 locus. The average H_o and H_e were 0.319 ± 0.237 and 0.343 ± 0.209 for 5/6 aiSNPs, respectively, and 0.435 ± 0.215 and 0.431 ± 0.217 for 5/6 aiSNP microhaplotypes, respectively. Prior to Bonferroni correction, no aiSNP microhaplotypes significantly deviated from HWE expectations. Before and after Bonferroni correction ($p < 3.25 \times 10^{-5}$), 381 and 22 aiSNP pairwise LDs were observed, respectively. No significant Bonferroni corrected LDs were observed between aiSNPs or aiSNP microhaplotypes on the same chromosome.

A minor increase in variation was observed for two piSNPs when considering flanking region sequences (Table 1 and Supplemental Table 6). Two piSNP microhaplotypes were defined as: mh16-MCR1 B and mh16-MCR1C [5]. The average piSNP microhaplotype H_o and H_e were 0.198 ± 0.213 and 0.190 ± 0.196, respectively. After Bonferroni correction ($p < 0.000549$), one (rs1805009/mh-16-MCR1B) significant pairwise LD was observed. It is not surprising to observe significant LD between rs1805009 and mh16-MCR1B due to their close physical proximity (~325 basepairs) within the MCR1 gene.

3.3. Bioinformatic concordance

Consistent with the ForenSeq™ UAS and STRbase [10,20], Wendt, et al. [10] did not include the DYS612 $[CCT]_a$ $[CTT]_b$ motif in length-based alleles. Based on Parson, et al. [16] recommendations and Novroski, et al. [3] population data, this region is now included for the length-based alleles at the DYS612 locus.

Compared to frequency data previously published by Wendt, et al. [10], the piSNP N29insA is discordant. The locus was reported as having no observed heterozygosity and yet the allele frequencies are 0.516 and 0.484 for the null and insertion alleles, respectively. By using STRait Razor v2s, a bioinformatic error was discovered. Manual confirmation of the locus in the ForenSeq™ UAS indicated that the error occurred when the "ForenSeqRunStatistics" XML files were offline. Thus, the values reported herein are a discordant due to an operation issue independent of analyses performed with the UAS. The N29insA frequencies reported here (1.00 null and 0.00 insertion) are correct.

4. Conclusion

Twenty-one human-identity markers have been identified in this study of the Yavapai population which contain some degree of flanking region variation, with a wide range of relative frequencies. A small portion of target autosomal STRs and iiSNPs exhibited flanking region variation at relatively high frequencies. The human identification marker set captured by the ForenSeq™ DNA Signature Prep Kit produced combined RMPs of 7.66×10^{-58} and

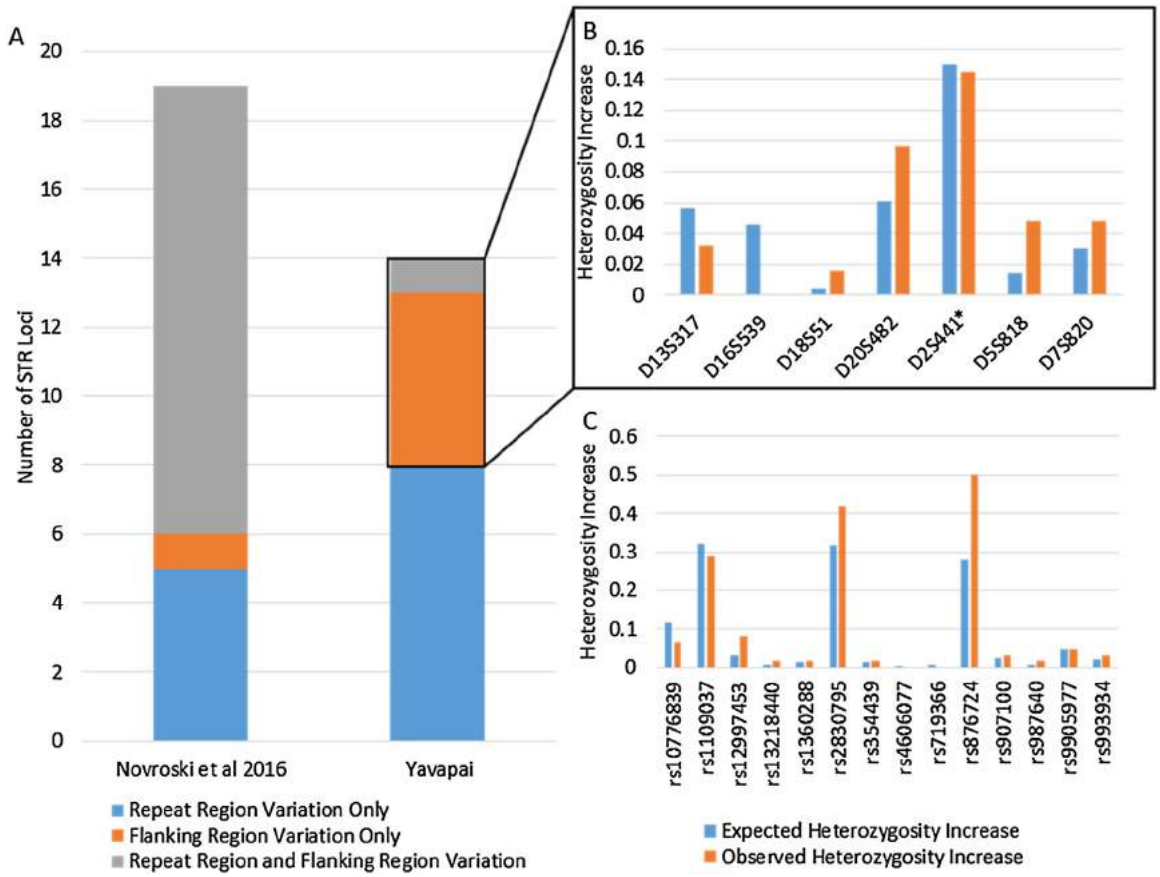

Fig. 1. Sequence variation of autosomal identity-informative markers. (A) Number of short tandem repeat (STR) loci in the Yavapai population and the Novroski, et al. [3] study exhibiting sequence variation; (B) Observed and expected heterozygosity increase of autosomal STR loci with flanking region variation; an asterisk indicates the locus with flanking region and repeat region variation; (C) Observed and expected heterozygosity increase of identity-informative single nucleotide polymorphism containing microhaplotypes.

5.49×10^{-63}, respectively, for a panel of 121 human identification markers (94 iiSNPs or MPS-based iiSNPs/iiSNP microhaplotypes plus 27 length-based or sequence-based autosomal STRs). With such low MPS-based RMPs, additional variation may not seem necessary, however, these results highlight that increased information can be obtained from the full amplicon.

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 Sreetharan Kanthaswamy and David Glenn Smith and a research grant to Kelly L McCulloh 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.2017.02.014.

References

[1] 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 (September) (2015) 118–130, doi:http://dx.doi.org/10.1016/j.fsigen.2015.06.005 Epub 2015 Jul 6. Review. PubMed PMID: 26197946.

[2] 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 (January) (2016) 20–29, doi:http://dx.doi.org/10.1016/j.fsigen.2015.09.009 Epub 2015 Sep 21 PubMed PMID: 26433485.

[3] N.M. Novroski, J.L. King, J.D. Churchill, L.H. Seah, B. Budowle, Characterization of genetic sequence variation of 58 STR loci in four major population groups, Forensic Sci. Int. Genet. 25 (September) (2016) 214–226, doi:http://dx.doi.org/10.1016/j.fsigen.2016.09.007 PubMedPMID:27697609.

[4] F.R. Wendt, D.H. Warshauer, X. Zeng, J.D. Churchill, N.M. Novroski, B. Song, J.L. King, B.L. LaRue, B. Budowle, Massively parallel sequencing of 68 insertion/deletion markers identifies novel microhaplotypes for utility in human identity testing, Forensic Sci. Int. Genet. 25 (September) (2016) 198–209, doi:http://dx.doi.org/10.1016/j.fsigen.2016.09.005 [Epubaheadofprint] PubMedPMID:27685342.

[5] K.K. Kidd, A.J. Pakstis, W.C. Speed, R. Lagacé, J. Chang, S. Wootton, E. Haigh, J.R. Kidd, Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics, Forensic Sci. Int. Genet. 12 (September) (2014) 215–224, doi:http://dx.doi.org/10.1016/j.fsigen.2014.06.014 PubMed PMID: 25038325.

[6] K.K. Kidd, W.C. Speed, Criteria for selecting microhaplotypes: mixture detection and deconvolution, Investig Genet. 6 (September (1)) (2015), doi:http://dx.doi.org/10.1186/s13323-014-0018-3 PubMed PMID :25750707; PubMed Central PMCID: PMC4351693.

[7] 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.

[8] 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.
[9] K.L. McCulloh, J. Ng, R.F. Oldt, J.A. Weise, J. Viray, B. Budowle, et al., The genetic structure of Native Americans in North American based on Globalfiler® STRs, J. Forensic Sci. (2016) submitted.
[10] F.R. Wendt, J.D. Churchill, N.M. Novroski, J.L. King, J. Ng, R.F. Oldt, K.L. McCulloh, J.A. Weise, D.G. Smith, S. Kanthaswamy, B. Budowle, Genetic analysis of the Yavapai Native American from west-central Arizona using the Illumina MiSeq FGx™ forensic genomics system, Forensic Sci. Int. Genet. 24 (September) (2016) 18–23, doi:http://dx.doi.org/10.1016/j.fsigen.2016.05.008 Epub 2016 May 17 PubMed PMID: 27243782.
[11] J.L. King, F.R. Wendt, N.M. Novroski, B. Budowle, STRait Razor v2s: advancing sequence-based allele reporting beyond STRs, Forensic Sci. Int. Genet. (2016) Submitted 2016 December.
[12] Lewis, Zaykin, Genetic Data Analysis Software, (1999) .
[13] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155 (June (2)) (2000) 945–959 PubMed PMID: 10835412 PubMed Central PMCID: PMC1461096.
[14] N.A. Rosenberg, J.K. Pritchard, J.L. Weber, H.M. Cann, K.K. Kidd, L.A. Zhivotovsky, M.W. Feldman, Genetic structure of human populations, Science 298 (December (5602)) (2002) 2381–2385 PMID: 12493913 PubMed.
[15] L. Porras-Hurtado, Y. Ruiz, C. Santos, C. Phillips, A. Carracedo, M.V. Lareu, An overview of STRUCTURE: applications, parameter settings, and supporting software, Front. Genet. 4 (May (98)) (2013), doi:http://dx.doi.org/10.3389/fgene.2013.00098 PubMed PMID: 23755071; PubMed Central PMCID: PMC3665925.
[16] W. Parson, D. Ballard, B. Budowle, J.M. Butler, K.B. Gettings, P. Gill, L. Gusmão, D. R. Hares, J.A. Irwin, J.L. King, Knijff Pd, N. Morling, M. Prinz, P.M. Schneider, C.V. Neste, S. Willuweit, C. 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 Sci. Int. Genet. 22 (May) (2016) 54–63, doi:http://dx.doi.org/10.1016/j.fsigen.2016.01.009 Epub 2016 Jan 21. PubMed PMID: 26844919.
[17] https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1& cad=rja&uact=8&ved=0ahUKEwjJ2p7-t9HQAhXIExoKHSkeB6wQFggbMAA&url=http%3A%2F%2Fmedicine.yale.edu%2Flab%2Fkidd%2Fpublications%2FMicrohapWebfile_1Oct_2015_237227_174718_29491. xlsx&usg=AFQjCNGnBIbD2IpGX-bdpYp5BXo1fs0w8g&sig2=5CRse-QRqv9gAirZejlbDw&bvm=bv.139782543,d.bGg.
[18] 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 PubMed PMID: 19083792.
[19] A.O. Tillmar, P. Mostad, T. Egeland, B. Lindblom, G. Holmlund, K. Montelius, Analysis of linkage and linkage disequilibrium for eight X-STR markers, Forensic Sci. Int. Genet. 3 (December (1)) (2008) 37–41, doi:http://dx.doi.org/10.1016/j.fsigen.2008.09.006 PubMed PMID: 19083865.
[20] National Institute of Justice STRBase, www.cstl.nist.gov/strbase/.