Int J Legal Med (2011) 125:559–563
DOI 10.1007/s00414-011-0568-3

ORIGINAL ARTICLE

Y-STR loci diversity in native Alaskan populations

Carey Davis · Jianye Ge · Abirami Chidambaram · Jonathan King · Meredith Turnbough · Michelle Collins · Orin Dym · Ranajit Chakraborty · Arthur J. Eisenberg · Bruce Budowle

Received: 3 January 2011 / Accepted: 14 February 2011 / Published online: 30 March 2011

Abstract Y chromosome short tandem repeat (Y-STR) loci are important genetic markers for forensic biological evidence analyses. However, paternal inheritance, reduced effective population size, and lack of independence between loci can reduce Y-STR diversity and may yield greater population substructure effects on a locus-by-locus basis compared with the autosomal STR loci. Population studies are necessary to assess the genetic variation of forensically relevant markers so that proper inferences can be made about the rarity of DNA profiles. This study examined 16 Y-STRs in three sampled populations of Native Americans from Alaska: Inupiat, Yupik, and Athabaskan. Population genetic and statistical issues addressed were: (1) the degree of diversity at locus and haplotype levels, (2) determination of the loci that contribute more so to haplotype diversity, and (3) the effects of population substructure on forensic statistical calculations of the rarity of a Y-STR profile. All three population samples were highly polymorphic at the haplotype level for the 16 Y-STR markers; however, the Native Americans demonstrated reduced genetic diversity compared with major US populations. The degree of substructure indicated that the three populations were related and admixed in terms of paternal lineage. The examination of more polymorphic loci may be needed to increase the power of discrimination of Y-STR systems in these populations.

Keywords Forensic DNA analysis · Y chromosome · STR · Haplotype · Native American

Electronic supplementary material The online version of this article (doi:10.1007/s00414-011-0568-3) contains supplementary material, which is available to authorized users.

C. Davis · J. Ge · M. Turnbough · R. Chakraborty ·
A. J. Eisenberg · B. Budowle (✉)
Department of Forensic and Investigative Genetics,
University of North Texas Health Science Center,
3500 Camp Bowie Blvd,
Fort Worth, TX 76107, USA
e-mail: Bruce.Budowle@unthsc.edu

J. Ge · J. King · M. Turnbough · R. Chakraborty ·
A. J. Eisenberg · B. Budowle
Institute of Investigative Genetics,
Department of Forensic and Investigative Genetics,
University of North Texas Health Science Center,
3500 Camp Bowie Blvd,
Fort Worth, TX 76107, USA

A. Chidambaram · M. Collins · O. Dym
State of Alaska, Department of Public Safety,
Scientific Crime Detection Laboratory,
5500 E Tudor Road,
Anchorage, AK 99507, USA

Introduction

The US population is composed of many different groups that may vary to some degree genetically. This variation should be considered when estimating the rarity of a DNA profile for paternity, kinship, and identity testing results by studying forensically relevant populations. The Native American Alaskan population, which accounts for 19% of Alaska's population [1], has not been substantially studied for genetic diversity, particularly for the loci used in DNA forensics. This population consists of several groups of people, such as the Inupiat, Yupik, Aleut, Tlingit, Haida, Tsimshian, Eyak, and North Athabaskan, which are distinguished primarily by their geographic and linguistic affiliations. The Yupik and Inupiat Eskimos and Athabaskan Indian populations broadly represent three major population groups of Native Alaskans. The Yupik primarily populate the Western and Southwestern region of Alaska and the Inupiat populate the Northwestern coast and the North Slope Borough region. The geographic boundary for the Yupik and Inupiat Eskimo groups is generally accepted to lie along the Norton Sound area and Seward Peninsula north of the town of Unalakleet. The Athabaskan Indian population primarily resides in the interior region of Alaska and is believed to have similar linguistic affiliations with those of some US Southwestern Native American populations such as the Apaches and Navajos.

Budowle et al. [2] previously described the genetic diversity of the 13 CODIS autosomal short tandem repeat (STR) loci in three Native American Alaskan populations: Athabaskans, Inupiats, and Yupiks. While all loci were highly polymorphic in the three Native Alaskan groups, genetic diversity was lower (by an extent of 4–15% in terms of average heterozygosity of the 13 CODIS STR loci) in the Native Alaskan populations compared with Caucasians and African Americans [2]. This reduced diversity is consistent with the ethnohistory of these population groups [3]. Additionally, the Athabaskans were found to be more closely related to Apaches and Navajos than the other two Native Alaskan groups that were studied [2].

Y chromosome short tandem repeat (Y-STRs) are short tandem repeat loci that reside specifically on the non-recombining portion of the Y-chromosome, and thus tend to be male specific. Therefore, Y-STR typing can be useful in paternal lineage studies and in typing male/female mixtures where the female component is substantially greater [4]. Although some data exist for Arctic region populations, such as the Siberian Eskimos and Canadian and Greenland Inuit populations [5–7], there are no data on the forensically relevant Y-STR loci in Native Alaskans. Because of the unique characteristics of the Y chromosome, the objective of this study was to determine genetic variation at 16 Y-STR loci for three of the major Native American Alaska populations.

Materials and methods

Samples

Buccal swabs or blood were collected from unrelated male convicted offenders who were required to provide a DNA sample according to Alaska Statute AS44.41.035. These samples were collected from each male participant by the Alaska Department of Corrections personnel and sent to the Scientific Crime Detection Laboratory where they were de-identified and assigned new code numbers. The sample populations consisted of 151 Inupiats, 150 Yupiks, and 153 Athabaskans. Ethnic affiliation was based on self-declaration at the time of sample collection.

The samples were then shipped to the Institute of Investigative Genetics at the University of North Texas Health Science Center (UNTHSC) in Fort Worth, TX, USA for further DNA typing. Except for population affinity, no personal information of the subjects is available to the researchers at the UNTHSC laboratory.

DNA extraction and quantification

The DNA was extracted using Qiagen's EZ1 DNA Investigator Kit on the EZ1 Advanced XL (Qiagen Inc., Valencia, CA, USA) following the manufacturer's protocol. Samples were eluted into 40 μL of 0.1 M Tris, 0.1 mM EDTA (TE^{-4}), and all extracted samples were stored at −20°C until analyzed.

The quantity of recovered DNA was determined using the Applied Biosystems Quantifiler® Human DNA Quantification Kit and the ABI 7500 Real-Time PCR System according to the manufacturer's instructions (Applied Biosystems, Foster City, CA, USA). Samples were normalized to 0.75 ng/μL.

Y-STR typing

The PCR amplification was performed with reagents contained in the AmpF*l*STR® Yfiler™ kit (Applied Biosystems) in a reduced reaction volume of 15 μL (14 μL of master mix and 1 μL of DNA). The master mix consisted of the following components for each sample: 5.8 μL of AmpF*l*STR® PCR Reaction Mix, 2.9 μL of AmpF*l*STR® Yfiler™ primers, 5.0 μL of ddH_2O, and 0.30 μL of 5 U/μL of AmpliTaq® Gold DNA polymerase (Applied Biosystems). Amplification was performed in an ABI PRISM® GeneAmp® 9700 Silver block Thermal Cycler (Applied Biosystems) using the 9600 emulation mode for 30 cycles. Prior to electrophoresis, 1 μL of the amplified product or allelic ladder and 0.5 μL of GeneScan™-500 LIZ® size standard (Applied Biosystems) were added to 8.5 μL of deionized Hi-Di™ formamide (Applied Biosystems), denatured at 95°C for 5 min, and placed on ice for 5 min. PCR products were separated and detected on an ABI PRISM® 3130 xl Genetic Analyzer (Applied Biosystems) following the manufacturer's recommendations. Samples were injected for 10 s at 3 kV and separated electrophoretically in performance optimized polymer (POP-4™; Applied Biosystems) using the HIDFragmentAnalysis36_POP4 Module (Applied Biosystems) and a 1,500 sec run time. The data were collected using the ABI PRISM® 3130 xl Genetic Analyzer Data Collection Software 3.0. Electrophoresis results were analyzed with GeneMapper® ID software v3.2 (Applied Biosystems). Allele peaks were called when the peak heights were equal to or greater than 50 relative fluorescence units.

Statistical analyses

Gene diversity at each locus, the number of haplotypes, and haplotype diversity were calculated using a program developed by Jianye Ge (UNTHSC, Ft Worth, TX, USA). Power of discrimination (PD; equivalently, haplotype diversity) was estimated as $1-\sum p_i^2$ where p_i is the observed relative haplotype frequency. Bias correction (i.e., multiplication of this expression by a factor of $N(N-1)^{-1}$) was not done because of the possibility that when each haplotype is observed once in a database of size N, bias-corrected estimate of PD would equal 1. F_{ST} was estimated following Weir and Cockerham [8], using haplotype data according to the logic described previously [9–12]. The coefficient of gene differentiation (G_{ST}) was estimated by the method described by Nei [13]. Genetic distances between populations were calculated by counting the sum of squared number of repeat differences between two haplotypes [14]. 3-D multidimensional scaling was plotted using Matlab [15] together with three major populations in the USA [10].

Data submission

This paper follows the recommendations of the ISFG on the use of Y-STRs in forensic analysis and the guidelines for publication of population data requested by the journal. Our institution passed the blind test (QC) offered by the YHRD and has been certified. The Y-STR population data have been submitted to the YHRD (www.yhrd.org) and the following YHRD accession numbers have been received after evaluation of the data: YA003681 (Alaska, USA [Inupiat]), YA003682 (Alaska, USA [Yupik]), YA003683 (Alaska, USA [Athabaskan]). All individual haplotypes are shown in Table S1 (electronic supplement data).

Table 1 PD per Y-STR marker per population

Locus	Inupiat	Yupik	Athabaskan
DYS456	0.5858	0.6207	0.7093
DYS389I	0.5078	0.5377	0.4792
DYS390	0.5532	0.5834	0.7068
DYS389II	0.6189	0.6324	0.7491
DYS458	0.7412	0.7633	0.7482
DYS19	0.4679	0.4582	0.7219
DYS385	0.9266	0.9411	0.9129
DYS393	0.6375	0.5997	0.4973
DYS391	0.3766	0.3036	0.5547
DYS439	0.6908	0.7641	0.7052
DYS635	0.5206	0.5425	0.7328
DYS392	0.7174	0.6588	0.7144
GATA H4	0.6832	0.6165	0.5805
DYS437	0.4119	0.4362	0.5598
DYS438	0.6272	0.5905	0.6681
DYS448	0.7274	0.7473	0.6843

Table 2 Haplotype data per population group

Population (sample size)	No. of distinct haplotypes	No. of population specific haplotypes	Haplotype diversity
Inupiat (N=151)	97	67	0.9805
Yupik (N=150)	109	77	0.9870
Athabaskan (N=153)	122	90	0.9903

Results and discussion

Population data

This study presents Y-STR population data on three Native Alaskan population groups: Inupiats, Yupiks, and Athabaskans. The DYS385 marker is the result of a tandem duplication and thus while technically 17 loci were analyzed the data herein will be treated as that from 16 loci (i.e., treating this duplicate locus as a single locus). The PD for each of the 16 Y-STR markers for the three Native American population groups is listed in Table 1. The PDs range from a low of 0.3036 for the DYS391 locus in Yupiks to a high of 0.9411 for the DYS385 locus in Yupiks. Due to the lack of biological independence among Y-STR markers, haplotype diversity is a better indicator of the power of

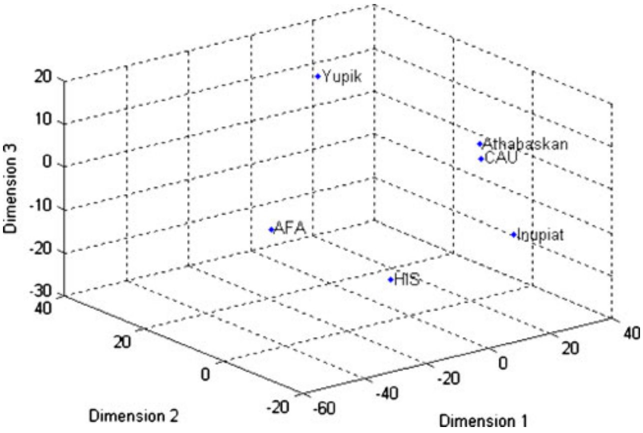

Fig. 1 3-D multidimensional scaling (*MDS*) plot of the six populations (i.e., three Native Alaskan, Texas Caucasian (*CAU*), Texas African American (*AFA*), and Texas Hispanic (*HIS*) populations [10]) using Matlab [15] based on pairwise genetic distances

Table 3 Number of shared distinct haplotypes among the population data sets

Sample size	Population	Inupiat	Yupik	Athabaskan	Afr. American	Caucasian
151	Inupiat					
150	Yupik	6				
153	Athabaskan	1	5			
1030	Afr. American	0	0	1		
1036	Caucasian	2	3	6	34	
1088	Hispanic	1	1	1	10	25

The African American, Caucasian, and Hispanic populations used were from the Budowle et al. [10] data set

discrimination of the Y-STR system than is individual locus diversity. The 16 marker haplotype diversities were 0.9805, 0.9870, 0.9903, for Inupiats, Yupiks, and Athabaskans, respectively (Table 2). Athabaskans had a higher PD than the other two Alaskan populations, which may be due to a higher degree of admixture of Athabaskan with the other major populations (Fig. 1). While the diversity of Y-STR haplotypes is high in all three Native American populations, it is approximately an order of magnitude lower compared with the three major populations in the USA [10, 16]. The proportions of distinct or population-specific haplotypes are also lower than observed in major populations [10, 16]. These observations are consistent with other studies [1, 16] and likely are due to the ethnohistory of these populations and to a lesser degree the smaller sample sizes in this study. For the same reasons, the proportions of the most frequent haplotypes (i.e., 9/151, 5/150, and 4/153 for Inupiat, Yupik, and Athabaskan, respectively) are higher than other major populations [10, 16].

Together with three major US populations (African American, Caucasian, and Hispanic) in [16], the numbers of shared haplotypes among the populations are shown in Table 3. Athabaskans share the most number of haplotypes with the major populations. These results are consistent with the genetic distance among the three native Alaskan populations and the three major US population groups (Fig. 1). Indeed, the Inupiat and Yupik are distant from the major populations while the Athabaskan population was closer to the Caucasian population. The maximum PD and accompanying F_{ST} and G_{ST} values for various numbers of markers comprising a Y-STR profile for the three Native Alaskan populations are listed in Table 4. With all 16 Y-STRs, the PD is 0.9951, which is higher than the PD of the individual populations. The F_{ST} and G_{ST} values are 0.0072 and 0.0092, respectively, which are higher than those observed for the major populations [10, 11, 16]. Therefore, the population correction for haplotype frequency calculation is not trivial in these Native Alaskan populations and should be included.

As recommended by the NRC II Report [17] and derived from Balding and Nichols [18], θ (which is derived from either F_{ST} or G_{ST} values) can be used to adjust for the effects of possible population substructure. Using the general theory, the unconditional frequency of a haplotype

Table 4 Maximum PD and accompanying F_{ST} and G_{ST} values for various numbers of markers comprising a Y-STR profile using the three Native Alaskan populations combined

Marker combination with maximum PD[a]	F_{ST}	G_{ST}	PD
6,	0.02777	0.02306	0.94861
6,9,	0.017	0.01579	0.97959
4,6,9,	0.01436	0.01401	0.98806
3,4,6,9,	0.01149	0.01209	0.99149
2,3,4,6,9,	0.01022	0.01123	0.99299
0,2,3,4,6,9,	0.00937	0.01067	0.99391
0,2,3,4,6,9,11,	0.00844	0.01004	0.99428
0,2,3,4,6,9,10,11,	0.0083	0.00995	0.99448
0,2,3,4,6,8,9,10,11,	0.00786	0.00966	0.99465
0,2,3,4,6,8,9,10,11,12,	0.00777	0.0096	0.9948
0,2,3,4,5,6,8,9,10,11,12,	0.00743	0.00937	0.99493
0,2,3,4,5,6,8,9,10,11,12,15,	0.00733	0.0093	0.995
0,1,2,3,4,5,6,8,9,10,11,12,15,	0.00725	0.00925	0.99504
0,1,2,3,4,5,6,7,8,9,10,11,12,15,	0.00718	0.0092	0.99508
0,1,2,3,4,5,6,7,8,9,10,11,12,13,15,	0.00715	0.00918	0.99509
0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,	0.00715	0.00918	0.99509

[a] 0=DYS456, 1=DYS389I, 2=DYS390, 3=DYS389II, 4=DYS458, 5=DYS19, 6=DYS385, 7=DYS393, 8=DYS391, 9=DYS439, 10=DYS635, 11=DYS392, 12=GATA_H4, 13=DYS437, 14=DYS438, 15=DYS448

(A_i), which is the count divided by the sample size, can be modified to obtain the conditional probability

$$\begin{aligned}\Pr(A_i|A_i) &= [p_i^2 + \theta p_i(1-p_i)]/p_i \\ &= p_i + \theta(1-p_i) \text{ or} \\ &= \theta + p_i(1-\theta)\end{aligned}$$

Therefore, the probability of observing a particular haplotype in an unrelated individual given that it is observed in another male is dependent on the value of θ. The F_{ST} and G_{ST} values for complete (and some degree of partial) haplotypes for major population groups have been so small as to have no impact on the upper-bound estimate of a Y-STR count proportion [10, 11, 16]. However, given the relatively high F_{ST} and G_{ST} values observed for the Native Alaskans, a θ correction is warranted based on the values described herein.

Lastly, we report the observation of a 15.1 allele at GATA-H4 locus in the Yupik sample B.17898.S. The 15.1 allele has not been observed in the YHRD (www.yhrd.org) or US Y-STR Database (http://www.usystrdatabase.org). This is an estimation of the allele size since the largest allele in the allelic ladder for this locus is a 13. The allele in the sample has yet to be sequenced.

Conclusion

This study examined the Y-STR diversity in three Native American populations and compared the results with that of three major US populations. The Native American data support that Y-STR haplotypes are highly polymorphic, although less than that of other forensically relevant US major populations. For profiles with a high number of Y-STR loci, the F_{ST} and G_{ST} of the Native Alaskan populations are close to 0.01, higher than those of the major populations. Consequently, population substructure correction is necessary for interpreting the Y-STR evidence for these lower diversity subpopulations. The addition of more polymorphic loci may increase the discrimination power of the analysis. Ballantyne et al. [19] examined 186 Y-STR loci and a subset of these may be beneficial in this scenario of reduced diversity populations to increase discrimination power.

References

1. Huntley K, Long J (2005) Gene flow across linguistic boundaries in Native North American populations. PNAS 102:1312–1317
2. Budowle B, Chidambaram A, Strickland L, Beheim C, Taft G, Chakraborty R (2002) Population studies on three Native Alaska population groups using STR loci. Forensic Sci Int 129:51–57
3. Burch ES Jr (1979) Indians and Eskimos in North Alaska, 1816–1977: a study in changing ethnic relations. Arct Anthropol 16 (2):123–151
4. Hammond HA, Jin L, Zhong Y, Caskey CK, Chakraborty R (1994) Evaluation of 13 short tandem repeat loci for use in personal identification application. Am J Hum Genet 55(1):175–189
5. Szathmary E (1993) Genetics of aboriginal North Americans. Evol Anthropol 1(6):202–220
6. Hallenberg C, Tomas C, Simonsen B, Morling N (2009) Y-chromosome STR haplotypes in males from Greenland. Forensic Sci Int Genet 3(4):e145–e146
7. Bosch E, Calafell F, Rosser Z, Nǿrby S, Lynnerup N, Hurles M, Jobling M (2003) High level of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosomal analysis. Hum Genet 112:353–363
8. Budowle B, Adamowicz M, Aranda X, Barna C, Chakraborty R, Eisenberg AJ, Frappier R, Gross AM, Lee HS, Milne S, Prinz M, Saldanha G, Krenke BE (2005) Twelve short tandem repeat loci Y chromosome haplotypes: genetic analysis on populations residing in North America. Forensic Sci Int 150(1):1–15
9. Budowle B, Ge J, Chakraborty R (2007) Basic principles for estimating the rarity of Y-STR haplotypes derived from forensic evidence. Eighteenth International Symposium on Human Identification 2007, Promega Corporation, Madison, Wisconsin. Available from: http://www.promega.com/ussymp18proc/default.htm
10. Budowle B, Ge J, Aranda X, Planz J, Eisenberg A, Chakraborty R (2009) Texas population substructure and its impact on estimating the rarity of Y-STR haplotypes from DNA evidence. J Forensic Sci 54(5):1016–1021
11. Budowle B, Ge J, Aranda X, Low J, Lai C, Yee WH, Law G, Tan WF, Chang YM, Perumal R, Keat PY, Mizuno N, Kasai K, Sekiguchi K, Chakraborty R (2009) The effects of Asian population substructure on Y-STR forensic analyses. Leg Med 11:64–69
12. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370
13. Nei M (1975) Molecular population genetics and evolution. North-Holland, Amsterdam, pp 149–150
14. Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462
15. Matlab. Available from: http://www.mathworks.com/help/toolbox/stats/cmdscale.html. Accessed on 10 Feb 2011
16. Ge J, Budowle B, Planz J, Eisenberg A, Ballantyne J, Chakraborty R (2010) US forensic Y chromosome short tandem repeats database. Leg Med 12(6):289–295
17. National Research Council (1996) The evaluation of forensic DNA evidence. National Academy Press, Washington
18. Balding DJ, Nichols RA (1994) DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 64:125–140
19. Ballantyne KN, Goedbloed M, Fang R, Schaap O, Lao O, Wollstein A, Choi Y, van Duijn K, Vermeulen M, Brauer S, Decorte R, Poetsch M, von Wurmb-Schwark N, de Knijff P, Labuda D, Vézina H, Knoblauch H, Lessig R, Roewer L, Ploski R, Dobosz T, Henke L, Henke J, Furtado MR, Kayser M (2010) Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. Am J Hum Genet 87(3):341–353