J Forensic Sci, May 2006, Vol. 51, No. 3
doi:10.1111/j.1556-4029.2006.00113.x
Available online at: www.blackwell-synergy.com

TECHNICAL NOTE

Alan J. Redd,[1†] Ph.D.; Veronica F. Chamberlain B.S.[1]; Veronica F. Kearney B.S.[1]; Daryn Stover,[1]Ph.D.; Tatiana Karafet,[1] Ph.D.; Karl Calderon[1]; Bruce Walsh,[2] Ph.D.; and Michael F. Hammer,[1,2] Ph.D.

Genetic Structure Among 38 Populations from the United States Based on 11 U.S. Core Y Chromosome STRs*

ABSTRACT: A DNA database consisting of the 11 Y chromosome short-tandem-repeat (Y-STR) recommended by the Scientific Working Group on DNA Analysis Methods is constructed for 2517 individuals from 38 populations in the United States. The population samples derive from five ethnic groups currently living in 10 states. A multidimensional scaling (MDS) plot places the populations into four discrete clusters (African Americans (AA), European Americans (EA), Hispanic Americans (HA), and Asian Americans (SA)) and one dispersed cluster of Native Americans. An analysis of molecular variance (AMOVA) indicates that a large proportion of the total genetic variance is partitioned among ethnic groups (24.8%), whereas only a small amount (1.5%) is found among-populations within ethnic groups. Separate AMOVA analyses within each ethnic group show that only the NA sample contains statistically significant among-population variation. Pair wise population differentiation tests do uncover heterogeneity among EA and among HA populations; however, this is due to only a single sample within each group. The analyses support the creation of AA, EA, HA, and Asian American databases in which samples from different geographic regions within the United States are pooled. We recommend that separate databases be constructed for different NA groups.

KEYWORDS: forensic science, population structure, Y chromosome DNA, short tandem repeats, African American, Hispanic American, European American, Native American, Asian American, DYS19, DYS385ab, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439

A key consideration for the proper scientific use of Y chromosome short-tandem-repeat (Y-STRs) in DNA forensics is the creation of an appropriate population database. A population database is necessary to estimate the probability that two or more unrelated males share the same Y-STR haplotype. To obtain an accurate estimate of a haplotype's frequency, the database should be large enough and represent the range of ethnic groups within a population (1). Otherwise, the frequency of a Y-STR haplotype of an individual whose ethnic group is not represented in the database is likely to be underestimated. Given that it is impossible to sample all individuals in a population, it is important to assess whether or not a population sample can be pooled with other populations from the same ethnic group that have been collected from different geographic regions (2). Because Y-chromosome haplotypes have been shown to exhibit large frequency differences among populations from different geographic regions (3,4) empirical studies are required to measure the proportion of variation within and among populations and ethnic groups (5) for forensic applications of Y-STRs in the United States.

[1]Division of Biotechnology, University of Arizona, Tucson, AZ 85721.
[2]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721.

*This project was presented at the NIJ DNA Grantees meeting in Washington, DC, June 27–29, 2005.

†Current address: Department of Anthropology, University of Kansas, Lawrence, KS 66045.

Received 24 Aug. 2005; and in revised form 4 Nov. 2005; accepted 6 Nov. 2005; published 5 April 2006

One such global database is the international Y-Chromosome Haplotype Reference Database (YHRD; http://www.yhrd.org/index.html), which consists of European, Asian, and U.S. population samples that have been typed for a set of nine Y-STRs (DYS19, DYS385ab, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393). This nine Y-STR panel comprises what is called the "minimal haplotype" (6,7). Kayser et al. (7) analyzed U.S. population structure with the minimal haplotype Y-STRs in 30 U.S. populations representing three ethnic groups: African American (AA), European American (EA), and Hispanic American (HA). Meanwhile, the Scientific Working Group on DNA Analysis Methods (SWGDAM) recommended adding two additional Y-STRs (DYS438 and DYS439) to the minimal haplotype, for a total of 11 loci to be analyzed in U.S. forensic work. To date, there is one study of U.S. population structure using the 11 U.S. Y-STRs, that of Budowle et al. (8) who examined population structure in 18 populations (14 U.S. populations and four Canadian populations). In this paper we further assess the extent of U.S. population structure with an analysis of 2517 samples representing 38 U.S. populations based on the 11 U.S. "core" Y-STRs.

Materials and Methods

DNA Samples

Samples for this study (Fig. 1 and Table 1) come from U.S. crime laboratories and include individuals from 10 states representing five ethnic groups including: AA ($n = 651$; 10 populations), EA ($n = 927$; 10 populations), HA ($n = 479$; nine

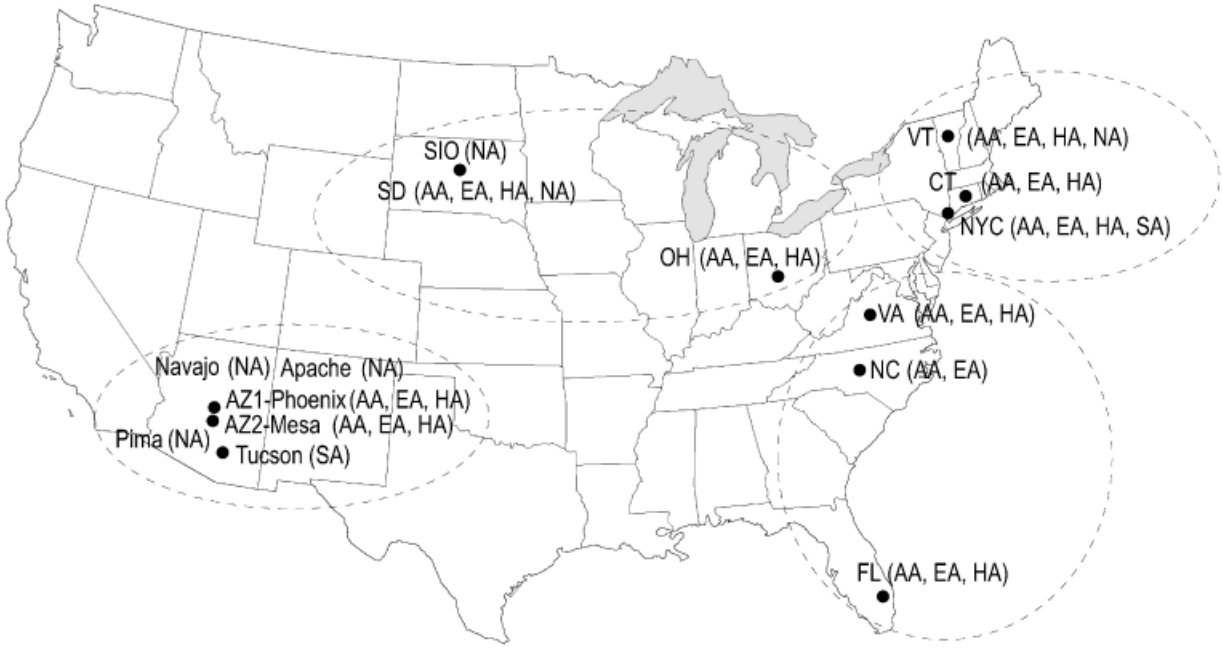

FIG. 1—*Map showing the approximate geographic positions of populations sampled in this study. The populations are grouped by ethnicity (AA, African American; EA, European American; HA, Hispanic American; NA, Native American; SA, Asian American) and by geography (dotted circles surround Southwest, Midwest, Northeast, and Southern samples). The Cheyenne sample is not shown.*

populations), Native Americans (NA) ($n = 398$; seven populations), and Asian Americans (SA) ($n = 62$; 2 populations). Samples were provided from the following colleagues: Jeffrey Ban (Virginia; Forensic Biology Section and DNA, Division of Forensic Science), Eric Buel (Vermont; Vermont Forensic Laboratory, State of Vermont, Department of Public Safety, Criminal Justice Services Division), Heather Miller Coyle (Connecticut; Forensic Science Laboratory State of Connecticut, Department of Public Safety, Criminal Justice Services Division), Cecelia Crouse (Florida; Serology/DNA section, Palm Beach Sheriff's Office Crime Laboratory), Dave Duplissa (Arizona-Phoenix; Arizona Department of Public Safety), Susan Narveson (Arizona-Phoenix; Arizona Department of Public Safety), Mark Nelson (North Carolina; Molecular Genetics Section North Carolina, State Bureau of Investigation, Department of Justice), Janice Nicklas (Vermont; Vermont Forensic Laboratory, State of Vermont, Department of Public Safety, Criminal Justice Services Division), Robin Pendergraft (North Carolina; Molecular Genetics Section North Carolina, State Bureau of Investigation, Department of Justice), Mecki Prinz (New York City; Department Forensic Biology, Office of Chief Medical Examiner, New York), Rex Riis (South Dakota; South Dakota Forensic Laboratory), Virginia Smart (Arizona-Mesa; Mesa Police Department Crime Laboratory), Mark Squibb (Ohio; Miami Valley Regional Crime Laboratory), and Russell Vossbrink (Arizona-Phoenix; Arizona Department of Public Safety). Samples were received as blood, blood stains, saliva stains, or purified DNA. Apache, Navajo, Pima, and Cheyenne samples have been described previously (9–12). All sampling protocols were approved by the Human Subjects Committee at the University of Arizona. Extraction and DNA quantification methods were previously described (13).

Multiplex PCR and DNA Typing

The U.S. set of core Y-STRs (DYS19, DYS385ab, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, and DYS439) were amplified using multiplex I and multiplex II as described in Redd et al. (13) with one modification: DYS438 primers from Butler et al. (14) were included in multiplex II. DYS389I was subtracted from DYS389II so that our analysis consisted of DYS389I and DYS389II-I. DYS385ab was analyzed by arbitrarily assigning the shorter allele to DYS385a and the longer allele to DYS385b. Positive and negative controls were included in all PCR reactions. All of the electropherogram data were entered into a database by two different individuals. Data were first checked and entered by either V. A. C. or V. A. K and then rechecked by A. J. R. Any unusual samples were reanalyzed.

Statistical Analyses

Haplotype diversity, the probability that two randomly chosen haplotypes are different in a sample, was calculated following Nei (15). Analysis of molecular variance (AMOVA) (16) analyses were carried out using the sum of the squared size differences (R_{ST}) using ARLEQUIN (Version 2.0; (17)). Populations were subdivided by ethnicity (AA, EA, HA, SA, and NA) and by geographic location, including the Southwest (Arizona and New Mexico), Midwest (South Dakota and Ohio), Northeast (Vermont, Connecticut, and New York), and South (Virginia, North Carolina, and Florida) (Fig. 1). Pairwise population differentiation (PPD) tests were performed in ARLEQUIN based on the sum of the squared number of repeat differences (R_{ST}). The significance of the AMOVA and PPD tests were assessed using 10,000 and 100,000 permutations, respectively. An R_{ST} genetic distance matrix that was generated using ARLEQUIN was used as input for a multidimensional scaling (MDS) analysis (18) using maximum likelihood estimation with the MULTISCALE program (J. O. Ramsay; ego.psych.mcgill.ca/pub/ramsay/multiscl). MDS is a metric scaling technique that allows one to visualize observed distances between samples in a simple two- or three-dimensional plot. Similar samples should be close together and dissimilar samples should be far apart. MDS iteratively compares the original distances with

TABLE 1—*Y-STR diversity in U.S. population samples.*

Ethnic Group Population	Sample Size	Number of Haplotypes	Discrimination Capacity (%)	Haplotype Diversity
African American (AA)	651	564	86.6	0.9994
Arizona-Phoenix (AZ1)	76	71	93.4	0.9982
Arizona-Mesa (AZ2)	52	49	94.2	0.9977
Connecticut (CT)	89	89	100.0	1.0000
Florida (FL)	20	20	100.0	1.0000
North Carolina (NC)	84	83	98.8	0.9997
New York City (NYC)	42	41	97.6	0.9988
Ohio (OH)	103	99	96.1	0.9992
South Dakota (SD)	57	57	100.0	1.0000
Virginia (VA)	77	71	92.2	0.9976
Vermont (VT)	51	49	96.1	0.9984
European American (EA)	927	664	71.6	0.9972
Arizona-Phoenix (AZ1)	56	54	96.4	0.9987
Arizona-Mesa (AZ2)	43	41	95.3	0.9978
Connecticut (CT)	85	76	89.4	0.9955
Florida (FL)	37	36	97.3	0.9985
North Carolina (NC)	87	81	93.1	0.9984
New York City (NYC)	42	42	100.0	1.0000
Ohio (OH)	99	87	87.9	0.9965
South Dakota (SD)	182	149	81.9	0.9968
Virginia (VA)	97	87	89.7	0.9970
Vermont (VT)	199	163	81.9	0.9958
Hispanic American (HA)	479	386	80.6	0.9981
Arizona-Phoenix (AZ1)	109	104	95.4	0.9992
Arizona-Mesa (AZ2)	47	44	93.6	0.9972
Connecticut (CT)	90	80	88.9	0.9973
Florida (FL)	20	18	90.0	0.9895
New York City (NYC)	38	32	84.2	0.9986
Ohio (OH)	24	24	100.0	1.0000
South Dakota (SD)	42	38	90.5	0.9954
Virginia (VA)	92	86	93.5	0.9981
Vermont (VT)	17	17	100.0	1.0000
Native American (NA)	398	259	65.1	0.9938
Apache (APA)	86	43	50.0	0.9436
Cheyenne (CHY)	29	27	93.1	0.9951
Navajo (NAV)	88	56	63.6	0.9804
Pima (PIM)	19	17	89.5	0.9883
South Dakota (SD)	112	91	81.3	0.9924
South Dakota-Sioux (SIO)	45	39	86.7	0.9909
Vermont (VT)	19	19	100.0	1.0000
Asian American (SA)	62	61	98.4	0.9995
Arizona-Tucson (AZ)	25	24	96.0	0.9967
New York City (NYC)	37	37	100.0	1.0000

Y-STR, Y chromosome short-tandem-repeat.

Euclidean distances computed from the plot, and then moves the samples around in the specified dimensional space in order to maximize the fit. Interpretations of the plot involve looking for meaningful clusters of samples.

Results

Haplotype Diversity

Haplotype diversity is uniformly high for most populations in this survey with NA having the lowest values of any ethnic group (Table 1). Average haplotype diversity at the population level ranges from a low of 0.984 for NA to a high of 0.999 for AA (data not shown). This trend is also reflected in the discrimination capacity (no. of haplotypes/no. of samples), which ranges from a low of 50% in the Apache to a high of 100% in several populations (Table 1). The SA sample has the highest discrimination capacity (98.4%) and haplotype diversity (0.9995) (Table 1). It is somewhat unexpected that diversity is higher in the SA vs. AA group given the higher diversity generally found in African populations (19) (although our Asian sample is small and this difference is not statistically significant). However, Budowle et al. (8) also found that their sample of 247 AA from four populations had slightly higher haplotype diversity than their SA sample, which was composed of 577 individuals deriving from five populations.

MDS Population Plot

Variation among the 38 U.S. populations can be seen in the maximum-likelihood MDS plot in Fig. 2. The fit between the original R_{ST} distance matrix and a genetic distance matrix derived from the plot is very high ($r = 0.98$) thus indicating that the MDS plot is a very good representation of the genetic distance matrix. Dotted circles are placed around populations from each ethnic group to illustrate that the population clusters correspond with five ethnic groups. The AA populations cluster to the top left of the plot, well separated from the other populations. Within the AA cluster, the FL AA population is on the far left of the cluster, while the OH AA population is on the far right of the cluster. The HA populations are found in close to the center in the plot. Within the HA cluster, the AZ2 (Mesa)-HA population is slightly separated to the right of the other HA populations. The EA populations form a tight cluster in the upper right of the plot adjacent to the HA cluster. Within the EA cluster, the NYC EA population is slightly removed from the other EA populations. The two SA populations cluster closely in the lower left quadrant of the plot and they are adjacent to most of the NA populations. In contrast, the NA populations are found across a large area of the MDS plot, they transect both the upper and lower right quadrants of the plot. In fact, the SD NA and SIO NA populations cluster very close to the HA cluster, and the VT NA population falls directly within the EA cluster.

AMOVA

Table 2 shows the results from analyses of molecular variance (AMOVA) based on the U.S. Y-STR loci. When populations were pooled into five ethnic groups most of the genetic variance (73.7%) is found within populations; a notable amount (24.8%) is found among ethnic groups; while only a small amount (1.5%) is found among populations within ethnic groups. Separate AMOVA analyses within each of the ethnic groups show that only the NA group contains significant among-population variation (9.5%; $p < 0.01$). Moreover, separate AMOVA analyses within ethnic groups that placed the population samples within their geographic locations, namely: Southwest, Midwest, Northeast, and South, indicate that only the NA populations contain significant among-group variation (11.1%). The AA, EA, HA, and SA ethnic groups do not contain significant substructure by geographic location (Southwest, Midwest, Northeast, and South) of the population samples.

Pairwise Differentiation Tests

Pairwise population differentiation tests mostly confirmed the patterns of genetic structure detected with the MDS and AMOVA analyses. There were no statistically significant differences among the 45 comparisons of pairs of AA samples (Table 3A). For EA samples, three of 45 comparisons (NYC-CT, NYC-VA, and NYC-NC) are found to be statistically significant at the $\alpha = 0.01$ level (Table 3B). Similarly, two of 36 comparisons between pairs of HA samples are statistically significant (Mesa-CT, Mesa-VA) (Table 3C). Notably, all of the statistically significant comparisons involved only a single sample within each group (NYC and Mesa).

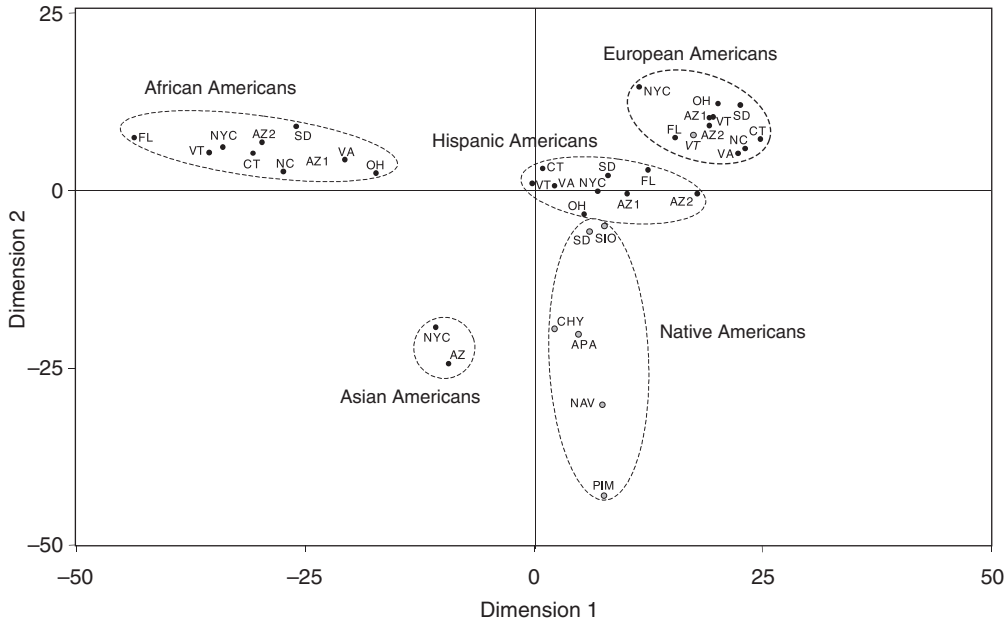

FIG. 2—*Multidimensional scaling (MDS) plot of 38 populations based on R_{ST} genetic distances. Population codes and symbols are the same as in Fig. 1. Native American samples are represented by gray dots, whereas all other samples are represented by black dots. Two and three letter codes are the same as in the Table 1.*

For NA samples, 13 of 21 comparisons are statistically significant (Table 3D), as expected based on the AMOVA and MDS results.

Discussion

This is the third large-scale study of the structure of Y-STR diversity in multiple U.S. populations (2,8), and only the second that we are aware of that screens the 11 core loci recommended by SWDAM (8). We find that levels of diversity based on the 11 U.S. core Y-STR loci (Table 1) are similar to those in Budowle et al.'s (8) 12 Y-STR analysis and consistently higher than those found by Kayser et al. (2) based on nine Y-STRs. For example, discrimination capacity for Kayser et al.'s (2) AA, EA, and HA samples were on average 8% lower than for our samples. Our survey differs from the other two published studies in that it examines multiple NA populations. The results indicate that NA populations have lower levels of Y chromosome diversity than other U.S. ethnic groups. This is reflected in a higher percentage of shared haplotypes and higher random match probabilities, both of which are important to take into consideration in forensic casework.

Similar to the results of Kayser et al. (2) we find that Y chromosomes are significantly differentiated among United States. ethnic groupings, but not among populations within ethnic groups from different geographic regions in the U.S. The proportion of among ethnic group variance reported here is almost identical to that found (24.9%) in the three ethnic groups (AA, EA, and HA) sampled by Kayser et al. (2), and higher (15.4%) than that reported in the five ethnic groups (AA, EA, HA, SA, and NA) sampled by Budowle et al. (8). Our AMOVA results differ somewhat from those in previous studies in that the proportion of variance among populations within groups is not significant for AA, EA, or HA, either when placing all geographic populations into a single ethnic grouping, or when subdividing populations within each ethnic group into geographic regions in the U.S. (Table 2). Kayser et al. (2) found that the very small proportion of among populations within groups variance in both their EA and HA samples (e.g., 1.8% and 2.6%, respectively) was statistically significant. The among-populations-within-groups variance (1.6%) for Budowle et al.'s (8) EA sample was statistically significant, but not for their HA sample (0.9%). Here, the only ethnic group with a statistically

TABLE 2—*Analysis of molecular variance (AMOVA) .*

Comparison	Number of Populations	Number of Groups	Among Groups	Among Populations Within Groups	Within Populations
African American (AA)	10	1	–	0.7	99.3
	10	4†	0.2	0.5	99.3
European American (EA)	10	1	–	0.2	99.8
	10	4†	0	0.2	99.8
Hispanic American (HA)	9	1	–	0.8	99.2
	9	4†	0.4	0.4	99.2
Native American (NA)	7	1	–	9.5*	90.5*
	7	3‡	11.1*	1.8*	87.1*
Asian American (SA)	2	na	–	0	100
All groups except NA	31	4	27.6*	0.4	72.0*
All groups	38	5	24.8*	1.5*	73.7*

*$p < 0.001$.
†Southwest, Midwest, Northeast, and South.
‡Southwest, Midwest, and East (Vermont).
na, not applicable.

TABLE 3A—p *values resulting from pairwise population differentiation tests on African American (AA) samples.*

	Phoenix	Mesa	CT	FL	NC	NYC	OH	SD	VA	VT
Arizona-Phoenix (AZ1)	*									
Arizona-Mesa (AZ2)	0.829	*								
CT	0.789	0.546	*							
FL	0.088	0.122	0.106	*						
NC	0.792	0.602	0.639	0.095	*					
NYC	0.554	0.382	0.875	0.173	0.373	*				
OH	0.213	0.112	0.048	0.011	0.125	0.056	*			
SD	0.542	0.705	0.333	0.065	0.211	0.271	0.103	*		
VA	0.481	0.273	0.140	0.019	0.269	0.119	0.868	0.218	*	
VT	0.360	0.544	0.562	0.343	0.500	0.644	0.028	0.145	0.067	*

CT, Connecticut; NC, North Carolina; FL, Florida; NYC, New York City; OH, Ohio; SD, South Dakota; VA, Virginia; VT, Vermont.

significant proportion of among populations within group variance is NAs, where we find 9.5% of the total variance partitioned among seven populations (tribes) (Table 2). Budowle et al. (8) found 3.0% of the total variance partitioned between their Navajo and Apache samples. The average among populations within groups variance in the three studies is 1.2%. When we remove NAs from our analysis, the among populations within groups variance is only 0.4% (not statistically significant; data not shown).

Despite the lack of significant differentiation among regional AA, EA, HA, and SA populations in AMOVA, when multiple differentiation tests are performed among all pairs of populations some comparisons between EA and HA populations are statistically significant. The question we face is whether these comparisons are significant by chance or as a result of true biological differences. Our results are very similar to those of Kayser et al. (2), who did not find different frequencies of Y-STR haplotypes among their AA samples, but did find heterogeneity within their EA and HA samples in pairwise population differentiation tests. As is the case here, this heterogeneity was attributed to a single sample within each group. Because they could not identify an obvious reason why either of the samples was an outlier, they concluded that their result reflected chance (2). We concur that a single outlier does not support a pattern of broad-scale geographic structuring (e.g., as observed for NA populations) and the combined results provide no compelling evidence for incorporating geographic structure within AA, EA, and HA Y-STR databases at present. One implication of these results is that independent databases can be combined for each of these ethnic groups. Still, it would be prudent to continue sampling from additional populations to further assess the structure of U.S. populations.

The extent to which we expect significant population structure within an ethnic group depends mainly on four factors: levels of subdivision in the ancestral source populations, the extent of nonrandom migration to the United States, migration rates among geographic regions after arrival in the United States, and the degree to which inter-ethnic admixture varies regionally. Kayser et al. (2) suggested that the lack of geographic heterogeneity

TABLE 3B—p *values resulting from pairwise population differentiation tests on European American (EA) samples.*

	Phoenix	Mesa	CT	FL	NC	NYC	OH	SD	VA	VT
Arizona-Phoenix (AZ1)	*									
Arizona-Mesa (AZ2)	0.968	*								
CT	0.541	0.805	*							
FL	0.797	0.813	0.260	*						
NC	0.436	0.636	0.943	0.474	*					
NYC	0.159	0.155	**0.004**	0.176	**0.004**	*				
OH	0.504	0.547	0.091	0.522	0.170	0.018	*			
SD	0.939	0.963	0.279	0.574	0.270	0.011	0.633	*		
VA	0.343	0.458	0.384	0.589	0.896	**0.007**	0.335	0.197	*	
VT	0.532	0.879	0.208	0.773	0.340	0.017	0.736	0.817	0.337	*

Bolded *p* values are ≤ 0.01.
CT, Connecticut; NC, North Carolina; FL, Florida; NYC, New York City; OH, Ohio; SD, South Dakota; VA, Virginia; VT, Vermont.

TABLE 3C—p *values resulting from pairwise population differentiation tests on Hispanic American (HA) samples.*

	Phoenix	Mesa	CT	FL	NYC	OH	SD	VA	VT
Arizona-Phoenix (AZ1)	*								
Arizona-Mesa (AZ2)	0.116	*							
CT	0.036	**0.006**	*						
FL	0.738	0.817	0.165	*					
NYC	0.479	0.142	0.219	0.479	*				
OH	0.804	0.084	0.282	0.512	0.630	*			
SD	0.657	0.070	0.256	0.399	0.370	0.467	*		
VA	0.138	**0.007**	0.686	0.215	0.271	0.693	0.333	*	
VT	0.234	0.030	0.734	0.313	0.417	0.645	0.263	0.743	*

Bolded *p* values are ≤ 0.01. CT, Connecticut; NC, North Carolina; FL, Florida; NYC, New York City; OH, Ohio; SD, South Dakota; VA, Virginia; VT, Vermont.

TABLE 3D—p *values resulting from pairwise population differentiation tests on Native American (NA) samples.*

	APA	CHY	NAV	PIM	SD	SIO	VT
APA	*						
CHY	0.547	*					
NAV	0.047	0.104	*				
PIM	**0.001**	**0.001**	**0.006**	*			
SD	**0.000**	0.022	**0.000**	**0.000**	*		
SIO	**0.001**	0.021	**0.000**	**0.000**	0.714	*	
VT	**0.000**	**0.000**	**0.000**	**0.000**	0.018	0.038	*

Bolded *p* values are ≤ 0.01.
APA, Apache; CHY, Cheyenne; NAV, Navajo; PIM, Pima; SD, South Dakota; SIO, Sioux; VT, Vermont.

among their AA samples may be a by-product of extensive migration from rural to urban areas during and after World War I. However, not enough is yet known about the structure of Y-STR haplotype variation among African source populations, or the extent of mixing among source populations in the process of forced migration to the United States. The finding of relatively high levels of population structure in NAs is not unexpected given a long history of small effective population sizes, endogamy, isolation, and founder effects (12). Perhaps it is more surprising that HA populations do not show stronger geographic structure given that the term Hispanic does not refer to a defined geographic region, but can refer to individuals of Mexican, Puerto Rican, Cuban, Central/South American, or other Spanish culture ancestry. In fact, HA populations are known to have differing degrees of Spanish, NA, and African ancestry in different U.S. regions (2,20). For example, Eastern Hispanics are expected to have more Afro-Caribbean ancestry than Hispanic populations from the Southwest, which are expected to have more NA ancestry (21). However, our Eastern HA populations do seem to cluster slightly closer to the AA populations than do the Southwest HA populations (Fig. 2).

In conclusion, the extremely consistent patterns of genetic structure observed in this study and previous studies (2,8) suggest that pooling samples from different geographic regions will not lead to strong biases in the estimation of Y-STR haplotype frequencies for AA, EA, HA, and SA populations. On the other hand, separate larger databases from NA subpopulations are needed to infer match probabilities for different tribal groups. Finally, the continued collection of core Y-STR data from additional populations is needed to ensure that we construct databases that most accurately reflect the structure of U.S. populations.

Acknowledgments

We thank John Butler for helpful comments. This research was supported by the National Institute of Justice Grant 2000-IJ-CX-K006 to M. F. H. We thank our colleagues from crime laboratories who supplied the U.S. population samples.

References

1. Lewontin R, Hartl D. Population genetics in forensic DNA typing. Science 1991;20:1745–50.
2. Kayser M, Brauer S, Schadlich H, Prinz M, Batzer M, Zimmerman P, et al. Y chromosome STR haplotypes and the genetic structure of U.S. populations of African, European, and Hispanic ancestry. Genome Res 2003;13:624–34.
3. Wilder J, Kingan S, Mobasher Z, Pilkington M, Hammer M. Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. Nat Genet 2004;36:1122–5.
4. Hammer MF, Zegura SL. The role of the Y chromosome in human evolutionary studies. Evol Anthropol 1996;5:116–34.
5. Zhivotovsky L, Ahmed S, Wang W, Bittles A. The forensic DNA implications of genetic differentiation between endogamous communities. Forensic Sci Int 2001;119:269–72.
6. Roewer L, Krawczak M, Willuweit S, Nagy M, Alves C, Amorim A, et al. Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. Forensic Sci Int 2001;118:106–13.
7. Kayser M, Brauer S, Willuweit S, Schadlich H, Batzer M, Zawacki J, et al. Online Y-chromosomal short tandem repeat haplotype reference database (YHRD) for U. W. populations. J Forensic Sci 2002;47:513–9.
8. Budowle B, Adamowicz M, Aranda XG, Barna C, Chakraborty R, Cheswick D, et al. Twelve short tandem repeat loci Y chromosome haplotypes: genetic analysis on populations residing in North America. Forensic Sci Int 2005;150:1–15.
9. Urbanek M, Goldman D, Long JC. The apportionment of dinucleotide repeat diversity in Native Americans and Europeans: a new approach to measuring gene identity reveals asymmetric patterns of divergence. Mol Biol Evol 1996;13:943–53.
10. Karafet T, Zegura SL, Vuturo-Brady J, Posukh O, Osipova L, Wiebe V, et al. Y chromosome markers and trans-bering strait dispersals. Am J Phys Anthropol 1997;102:301–14.
11. Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, et al. Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. Am J Hum Genet 1999;64:817–31.
12. Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF. High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. Mol Biol Evol 2004;21:164–75.
13. Redd A, Agellon A, Kearney V, Contreras V, Karafet T, Park H, et al. Forensic value of 14 novel STRs on the human Y chromosome. Forensic Sci Int 2002;130:97–111.
14. Butler J, Schoske R, Vallone P, Kline M, Redd A, Hammer M. A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. Forensic Sci Int 2002;129:10–24.
15. Nei M. Molecular evolutionary genetics. New York: Columbia University Press; 1987.
16. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 1992;131:479–91.
17. Schneider S, Kueffer J-M, Roessli D, Excoffier L. ARLEQUIN: a software for population genetic analysis. Geneva: Genetics and Biometry Laboratory, Stanford University; 2000.
18. Kruskal JB. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 1964;29:1–27.
19. Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, et al. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet 2000;66:979–88.
20. Merriwether D, Huston S, Iyengar S, Hamman R, Norris J, Shetterly S, et al. Mitochondrial versus nuclear admixture estimates demonstrate a past history of directional mating. Am J Phys Anthropol 1997;102:153–9.
21. Chakraborty BM, Fernandez-Esquer ME, Chakraborty R. Is being Hispanic a risk factor for non-insulin dependent diabetes mellitus (NIDDM)? Ethnic Dis 1999;9:278–83.

Additional information and reprint requests:
Michael F. Hammer, Ph.D.
Division of Biotechnology
University of Arizona
Tucson, AZ 85721
E-mail: mfh@email.arizona.edu