

Ricco, Emily M., Y-chromosome Heterogeneity in Three Native North American Populations. Master of Science (Forensic Genetics), May, 2013, 25 pp., 5 tables, 5 figures, 32 references.

Y-chromosome STR haplotype databases for Sioux, native Alaskan, and general Native American populations were used to predict Y-haplogroup designations. A total of 156 Sioux collected from seven geographic regions of South Dakota, 448 native Alaskan collected from three different tribes of Alaska, and 105 undefined Native American Y-STR haplotypes were available for analysis. Haplogroups were defined using prediction software that uses a Bayesian model to estimate haplogroup probabilities from STR-haplotype data. The frequency distribution of Y-haplogroups within these Native American populations was calculated from the resulting probabilities to determine the geographical proportions of non-Native American haplogroups. Inter-population heterogeneity was examined with a comparison of contributing haplogroups between the three groups of Native Americans.

The results establish the presence of population substructuring within the Native American groups investigated here. This can have implications for the interpretation of Y-STR data in forensic casework, particularly in calculating the rarity of a profile from representative population databases. Accounting for substructuring in genetically heterogeneous populations would facilitate the collection of truly representative samples to be included in forensically relevant databases.

Y-CHROMOSOME HETEROGENEITY IN THREE NATIVE NORTH AMERICAN
POPULATIONS

THESIS

Presented to the Graduate Council of the
Graduate School of Biomedical Sciences

University of North Texas

Health Science Center at Fort Worth

in Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

By

Emily M. Ricco, B.A.

Fort Worth, Texas

May 2013

ACKNOWLEDGMENTS

I'd like to thank my major professor, Dr. Ranajit Chakraborty, for his invaluable guidance throughout the semester. I'd like to thank members of my committee, Dr. John Planz, Dr. Robert Barber, and Dr. Andras Lacko, for their input. My thanks go to Stacey Jo Smith for providing her data. I'd also like to thank my friends and fellow students for their continued support over the last two years. Lastly, I owe a great deal of gratitude to my parents and family for providing a strong emotional and financial foundation for me. Your love and encouragement has been a guiding force.

TABLE OF CONTENTS

LIST OF FIGURES.....	v
LIST OF TABLES	vi
CHAPTER ONE: INTRODUCTION.....	1
CHAPTER TWO: MATERIALS AND METHODS	
2.1 Sample Collection and Preparation	7
2.2 Data Collection.....	8
2.3 Data Analysis	8
2.3.1 Y-haplogroup distribution	8
2.3.2 STRUCTURE Analysis	9
2.3.3 NETWORK Analysis	9
CHAPTER THREE: RESULTS	10
3.1 Distribution of Y-haplogroups	15
3.2 STRUCTURE	17
3.3 NETWORK.....	18

CHAPTER FOUR: DISCUSSION	21
LIST OF REFERENCES	23

LIST OF FIGURES

1. Tribal Lands Map of South Dakota.....	2
2. Global Distribution of Y Haplogroups.....	5
3. Haplogroup Proportions of African (E), Native American (Q), European (R), and All Others.....	16
4. Triangle Plot of Individual Membership Coefficients	17
5. Median-joining Network Using 11 Y-STR Loci Relating Haplotypes of Individuals from 3 Populations.....	20

LIST OF TABLES

1. Areas of Use of Y-chromosome Testing.....	4
2. Allele Frequency and PD Estimates by Locus per Population.....	10
3. Number of Haplotypes at Frequencies of 1-15 Copies	14
4. Number of Shared Haplotypes Among Population Data Sets	15
5. Average Proportion of Membership of Each Population in Each of the 3 Clusters	18

CHAPTER ONE: INTRODUCTION

Native American populations of North America represent diverse groups separated geographically, linguistically, and genetically. It is widely accepted that North America was initially peopled via migratory waves from Siberia, although the number and timing of immigration events as well as the route(s) taken are disputed (1-3). Origins and movements of modern-day native populations in the Americas rely, in part, on information gleaned from genetic studies using autosomal, mitochondrial and/or Y-chromosomal DNA data (4-6). Despite their common ancestral origin, the Sioux and Native Alaskan populations investigated in this study have distinct genetic histories that have been uniquely shaped by more recent inter- and intra-ethnic interactions.

The Sioux originally inhabited lands throughout the Missouri valley and by the 1700s many had migrated to present-day South Dakota (7). Today, the South Dakota Sioux represent nine federally recognized tribes, seven of which are included in this study – Cheyenne River, Lower Brule, Oglala (who reside in the Pine Ridge reservation), Rosebud, Sisseton, Standing Rock, and Yankton (Figure 1)(8).

The three Native Alaskan groups in this study – the Inupiat, Yupik, and Athabaskans – originate from three distinct geographical regions in the state of Alaska. The Inupiat are primarily situated along the Northwestern coast, the Yupik in the Southwest, and the Athabaskan population occupies a large interior region (9).

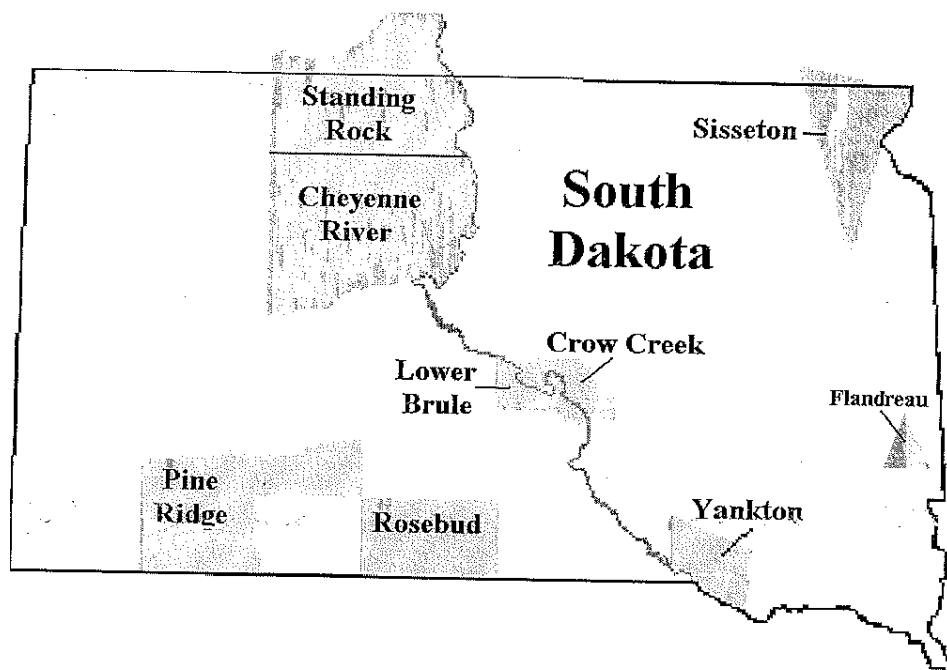


Figure 1. Tribal lands map of South Dakota (adapted from 8).

Y-chromosomal tests have a broad range of applications in the forensic and anthropological fields, from sexual assault casework and missing persons investigations to studies about human evolution, population migration, and paternal lineages (Table 1)(10). Two common analysis methods utilize Y-chromosome short tandem repeats (Y-STRs) and single nucleotide polymorphisms (Y-SNPs). The unique nature of Y-chromosome data comes from its non-recombinant nature. Haplotypes (composite profiles from a panel of STR or SNP loci) are, therefore, inherited as a unit from father to son and can be tracked across generations. Microsatellite regions of the Y-chromosome genome are highly variable and panels of Y-STR markers have been developed providing high discriminatory power for use in forensic cases (11). Forensic statistical analyses involve the estimation of Y-STR haplotype frequencies in the relevant population(s) through the use of haplotype databases. The presence of substructure can

affect the distribution of haplotype frequencies among populations and must be taken into account when compiling databases (12).

Analysis of Y-haplogroups has become an area of increasing interest. A Y-haplogroup represents a group of Y-chromosome haplotypes related by descent from lineages carrying specific mutations at defined SNP sites. Haplogroups are typically determined through the analysis of diagnostic SNP markers, although this is not always manageable or convenient in terms of time-to-result (13). Furthermore, definition of haplogroups, by itself, does not provide information related to the rarity of any specific haplotype in a population. Consequently, a method has been developed for predicting Y-haplogroups based on routinely used Y-STR markers in the absence of Y-SNP data (13-15). The current model is based on Bayesian principles of allele frequency variation. Y-chromosome STR data can be used to determine the probability that a given haplotype belongs to a certain haplogroup (15).

The Y-haplogroup predicting software that was employed in this study has recently come under scrutiny with questions arising as to its efficiency and reliability. Users are advised to exercise caution given the findings, notably a high probability of error in assigning haplogroups and a bias towards haplogroup R (16). However the predictor does not unequivocally assign a haplogroup; rather, it provides the probability of a given haplotype falling into any number of twenty-one haplogroups. Atthey recognizes the limitations of this program citing the unavailability of an adequate number of haplotypes from which these probabilities are calculated. Additionally, substructure cannot be taken into account for any affected haplogroups due to insufficient data (13, 15).

Use	Advantage
Forensic casework on sexual assault evidence	Male-specific amplification (can avoid differential extraction to separate sperm and epithelial cells)
Paternity testing	Male children can be tied to fathers in motherless paternity cases
Missing persons investigations	Patrilineal male relatives may be used for reference samples
Human migration and evolutionary studies	Lack of recombination enables comparison of male individuals separated by large periods of time
Historical and genealogical research	Surnames usually retained by males; can make links where paper trail is limited

Table 1. Areas of use in Y-chromosome testing (adapted from 10).

There have been numerous studies addressing haplogroup lineages, evaluating genetic distances between populations, and the resulting theories about early human migrations (17-19). Haplogroups often reflect different ethnic and geographical distributions based on the historical and cultural context of the population(s) in question. An estimated distribution of founder Y-haplogroups based upon Y-SNP haplotypes from around the world has been developed (Figure 2)(20). Y-SNP data for the three North American populations, the Navajo, Cheyenne, and Mixtecs, were taken from Karafet *et al.* (1) based on their geographical distribution and inclusion of relevant Y-SNP markers (20). Note that in North, Central, and South America the predominant haplogroup represented is Q with variable proportions of P, R, and C among others primarily Asian and European in origin.

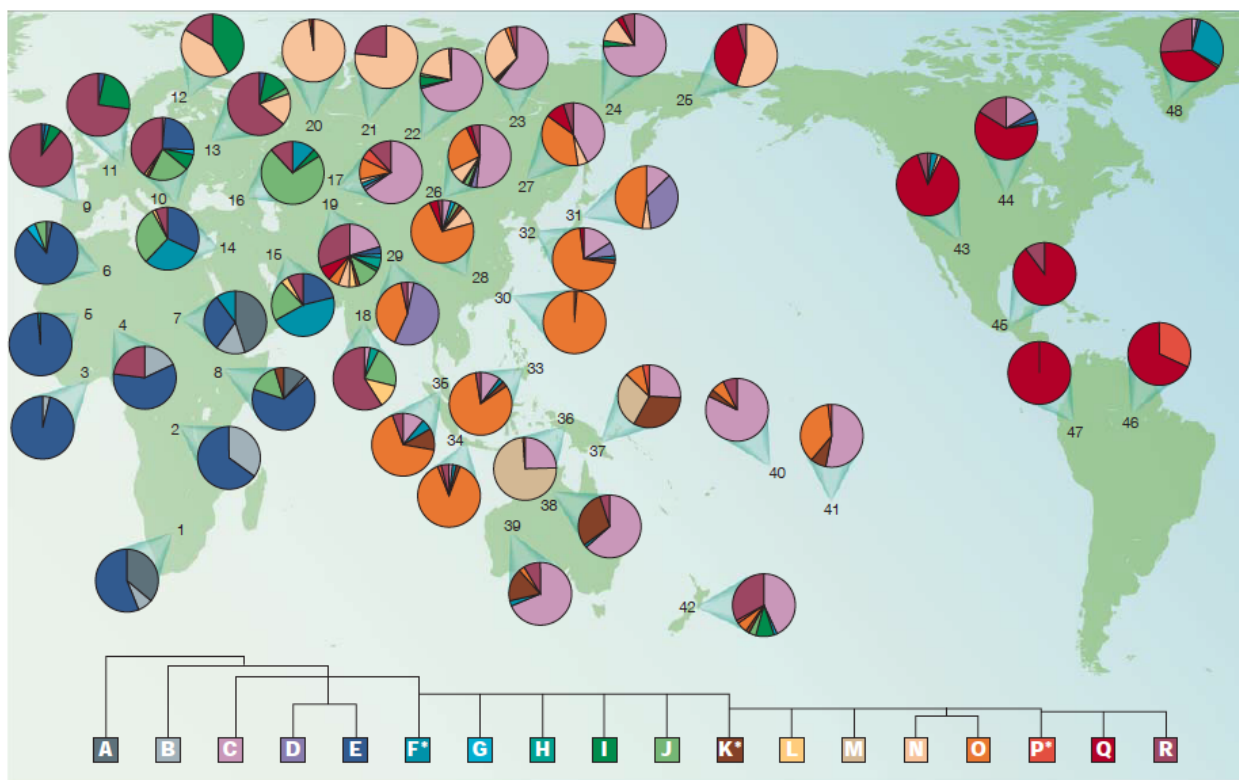


Figure 2. Global distribution of Y haplogroups (adapted from 20).

Determining the proportion of non-Native American haplogroups within populations is necessary in evaluating the presence and degree of admixture. Analysis of substructured populations can identify subgroups with distinctive allele frequency compositions (21). Heterogeneity caused by inter-ethnic admixture and substructure can affect the interpretation of Y-STR data and estimation of haplotype frequencies. Hammer *et al.* stressed the importance of careful selection of samples that are truly representative of the population in question, particularly in forensic casework (22). Assessing the degree of heterogeneity among Native American haplogroups can be useful in forensic cases when dealing with an admixed population, which is generally the case for all Native American populations encountered. Population substructure studies can determine whether or not region-specific forensic DNA databases need to be constructed based on the level of observed heterogeneity within Native

American populations and whether or not theta (θ) corrections need to be made in statistical calculations (12, 23). Further, presence of substantial proportions of non-Native American haplogroups in any Native American samples of individuals precludes evaluation of rarity of any specific target haplotype found in a Native American only by considering databases of Native American samples.

My hypothesis is that the Sioux, Native Alaskan, and unclassified Native American population samples show some level of inter- and intra-population Y-haplogroup variation. Specifically, I expect these Native American populations to have a considerable amount of Caucasian admixture (i.e. a higher proportion of European haplogroups in comparison to other non-Native American haplogroups). Y-STR haplotype data from previous studies were used to determine the distribution of Y-haplogroups in the three sample populations. The proportion of non-Native American haplogroups in each population was determined and the degree of Y-chromosome heterogeneity among the three populations was evaluated using genetic data analyses.

CHAPTER TWO: MATERIALS AND METHODS

2.1 Sample Collection and Preparation

Samples for the construction of the Sioux Y-STR dataset were collected from the South Dakota Violent Offender Database (8). A total of 156 individuals were typed at nineteen loci yielding 129 distinct haplotypes from seven geographic regions of the state of South Dakota. Tribal designation was determined by the individual's place of birth within a specific region/reservation of South Dakota.

Samples for the construction of the native Alaskan Y-STR dataset were collected mainly from offenders, as required by Alaskan Statute AS44.41.035, with some samples having been donated by volunteers (9). A total of 448 individuals were typed at sixteen Y-STR loci. The sample set is comprised of 150 Inupiat, 146 Yupik, and 152 Athabaskans. Ethnic affiliation was self-described.

Samples for the construction of the ABI Y-STR database were previously typed using the AmpFℓSTR® Yfiler™ PCR Identification kit. A total of 105 Native American haplotypes are included without specification of the tribal or geographic origins of the sampled individuals (25).

2.2 Data Collection

The Y-haplotypes from each dataset were used as input variables into Athey's Haplogroup Predictor to obtain the probability of each haplotype examined being assigned to a specific haplogroup (13-15).

2.3 Data Analysis

Allele frequencies at each locus, number of haplotypes, and number of shared haplotypes between populations were calculated using Arlequin v.3.5 (26). The probability of discrimination (PD) for each locus was calculated as the complement of sum of squares of allele frequencies.

2.3.1 Y-haplogroup distribution

To evaluate the distribution of haplogroups, probabilities for each of the twenty-one haplogroups, calculated by the Haplogroup Predictor, were combined across all samples in each dataset. Comparison of total haplogroup probabilities between the ABI, Native Alaskan, and Sioux datasets was carried out using a permutation-based test of RxC contingency table data (27). This same comparison was performed between the seven subgroups of the Sioux population and between the three subgroups of the Native Alaskan population. Probabilities for any subclades of haplogroups E, Q, and R were then combined. All remaining haplogroup probabilities were totaled and the proportion of each of these four groupings was determined for all three populations. Comparison of the proportion of haplogroups E, Q, R, and all others was carried out using the same permutation-based test of RxC contingency table data.

2.3.2 *STRUCTURE Analysis*

To infer population structure from the Y-haplotype data, the software package STRUCTURE v.2.3.4 was run using both the correlated allele frequencies and admixture models with a burn-in of 10,000 and 50,000 iterations (28-29). The program assigns individuals to populations or clusters (K), which are defined by their allele frequency compositions across all loci used in the analysis, by estimation of a membership coefficient/probability (Q). Under the admixture model, individuals showing some level of admixture can be assigned to two or more clusters. Given the results of the permutation-based tests of RxC contingency table data performed on the three Native American populations in this study, K was set to 3.

2.3.3 *NETWORK Analysis*

A Median-joining (MJ) network was constructed from the Y-STR data using NETWORK v.4.611 to further elucidate the presence of population structure and the relationship between haplotypes of individuals from the three sample populations (30-31). The loci used in the analysis were those that were typed in all three populations for a total of 12 loci (loci 389I and 389II were omitted from analysis per the developer's instructions)(30). Weights and epsilon parameters were set to their default values of 10 and 0, respectively.

CHAPTER THREE: RESULTS

Allele frequency estimates per locus for all population data sets are provided in Table 2, along with their locus-specific PD values.

Locus	Allele	Sioux	Native Alaskan	ABI
DYS19	12	0.01923	0.02232	0.00952
	13	0.12821	0.55134	0.14286
	14	0.64103	0.27232	0.57143
	15	0.16026	0.11384	0.14286
	16	0.05128	0.02232	0.09524
	17	0	0.01786	0.0381
	PD	0.54748	0.60895	0.62802
DYS389I	11	0	0.00223	0
	12	0.33333	0.0692	0.14286
	13	0.5641	0.46205	0.64762
	14	0.10256	0.45536	0.20952
	15	0	0.01116	0
	PD	0.56377	0.57552	0.52125
DYS89II	27	0.00641	0.00223	0.01905
	28	0.27564	0.07366	0.08571
	29	0.41667	0.19866	0.44762
	30	0.21795	0.47545	0.30476
	31	0.04487	0.20536	0.09524
	32	0.03846	0.04018	0.0381
	33	0	0.00446	0.00952
	PD	0.70389	0.68678	0.69505

Table 2. Allele frequency and PD estimates for all loci per population.

Locus	Allele	Sioux	Native Alaskan	ABI
DYS390	20	0	0.00223	0
	21	0.02564	0.00893	0.08571
	22	0.03846	0.07366	0.08571
	23	0.33333	0.23214	0.24762
	24	0.33333	0.53125	0.45714
	25	0.25641	0.125	0.11429
	26	0.01282	0.02679	0.00952
PD		0.71431	0.64347	0.70861
DYS391	9	0.12179	0.04911	0.01923
	10	0.59615	0.72098	0.60577
	11	0.26923	0.21652	0.36538
	12	0.01282	0.01116	0.00962
	13	0	0.00223	0
PD		0.56071	0.43173	0.50392
DYS392	9	0	0	0.00952
	11	0.28846	0.16404	0.28571
	12	0.02564	0.01573	0.05714
	13	0.25641	0.20225	0.42857
	14	0.37179	0.38427	0.1619
	15	0.03846	0.21798	0.02857
	16	0.01923	0.01348	0.02857
PD		0.71489	0.73823	0.71026
DYS393	10	0	0.00223	0
	12	0.19231	0.08482	0.05769
	13	0.52564	0.47098	0.75
	14	0.25641	0.37723	0.15385
	15	0.02564	0.0625	0.03846
	16	0	0.00223	0
PD		0.62432	0.62616	0.41299
DYS434	13	0	0	0.0381
	14	0.63462	0.32143	0.34286
	15	0.29487	0.61384	0.50476
	16	0.0641	0.06473	0.11429
	17	0.00641	0	0
PD		0.50943	0.51685	0.61905

Table 2 continued.

Locus	Allele	Sioux	Native Alaskan	ABI
DYS437	8	0.01923	0	0
	9	0.25641	0.00893	0.02857
	10	0.42308	0.31473	0.2
	11	0.30128	0.45982	0.21905
	12	0	0.21429	0.50476
	13	0	0.00223	0.04762
PD		0.6684	0.64494	0.66044
DYS438	8	0	0.00223	0
	10	0.04487	0.05134	0.04762
	11	0.39744	0.375	0.25714
	12	0.48077	0.26786	0.51429
	13	0.05128	0.23438	0.1619
	14	0.02564	0.0558	0.01905
	15	0	0.01116	0
PD		0.60951	0.72844	0.6467
YGATH4	10	0.03205	0.21973	0.01923
	11	0.30769	0.4148	0.34615
	12	0.57692	0.33408	0.51923
	13	0.07051	0.03139	0.11538
	14	0.01282	0	0
PD		0.56998	0.66856	0.60269
DYS635	20	0.01923	0.00893	0.01905
	21	0.10256	0.0558	0.12381
	22	0.41026	0.52679	0.2381
	23	0.38462	0.29464	0.51429
	24	0.05128	0.07812	0.05714
	25	0.03205	0.03125	0.04762
	26	0	0.00446	0
PD		0.67353	0.62679	0.66392

Table 2 continued.

Locus	Allele	Sioux	Native Alaskan	ABI
DYS385a/b	9,14	0	0.00223	0.00952
	10,13	0	0.00223	0
	10,14	0.00641	0.01116	0.01905
	11,11	0.00641	0.00223	0.01905
	11,12	0.00641	0.01116	0.01905
	11,13	0.06410	0.04018	0.06667
	11,13.2	0	0	0.00952
	11,14	0.18590	0.15625	0.32381
	11,15	0.01923	0.03125	0.05714
	11,16	0.01282	0.00446	0
	11,20	0	0.00893	0
	12,12	0	0.01563	0
	12,13	0.00641	0.00223	0.02857
	12,14	0.02564	0.00446	0.02857
	12,15	0.01282	0.02679	0.00952
	12,16	0	0.00670	0
	12,17	0	0.00223	0.00952
	12,19	0.01282	0.02902	0
	12,20	0.01282	0.00446	0
	12,21	0.02564		0
	13,13	0	0.02009	0
	13,14	0.03205	0.01563	0.04762
	13,15	0.01923	0.00223	0
	13,16	0.01923	0.00446	0.01905
	13,17	0.01923	0.02232	0
	13,18	0	0	0.00952
	13,19	0.03846	0.03348	0.00952
	13,20	0.00641	0.04688	0
	13,21	0	0.02902	0
	13,23	0	0.00223	0
	14,14	0.01282	0.02902	0.03810
	14,15	0.02564	0.02679	0.03810
	14,16	0.02564	0.00223	0.01905
PD		0.91535	0.94877	0.87256

Table 2 continued.

A total of 121 unique haplotypes were characterized from the 156 Sioux individuals, the majority of which are not shared within the population, with 102 haplotypes (or 84.3%) occurring only once (Table 3). Native Alaskan haplotypes exhibited a greater degree of intra-population sharing with 280 unique haplotypes, of which 204 (or 72.86%) are singly represented. Among the ABI individuals, 98 out of the 101 (97.03%) unique haplotypes were present as a single copy indicating that this population has the lowest incidence of haplotype sharing. The Native Alaskan and ABI populations displayed the only incidences of inter-population haplotype sharing (Table 4).

No. of occurrences	No. of haplotypes		
	Sioux	Native Alaskan	ABI
1	102	204	98
2	15	42	2
3	3	13	1
4	0	7	0
5	0	5	0
6	0	3	0
7	0	3	0
8	0	0	0
9	0	2	0
10	0	0	0
11	0	1	0
12	0	0	0
13	0	0	0
14	0	0	0
15	1	0	0
No. of distinct haplotypes			
	121	280	101

Table 3. Number of haplotypes at frequencies of 1 to 15 copies per population data set.

Sample Size	Population	Sioux	Native Alaskan
156	Sioux		
448	Native Alaskan	0	
105	ABI	0	6

Table 4. Number of shared haplotypes among the population data sets.

3.1 Distribution of Y-chromosome haplogroups

The distribution of the twenty-one Y-haplogroups as well as the distribution of the collapsed haplogroups was significantly different between the ABI, Native Alaskan, and Sioux populations ($p < 10^{-4}$) and between the three subgroups of the Native Alaskan population ($p < 10^{-4}$). Specifically, the distribution of Q, R, and all other haplogroups was shown to be significantly different between all three populations. Haplogroup distribution between the seven subgroups of the Sioux population was not significantly different for the twenty-one haplogroup comparison ($p = 0.0856$) or the collapsed haplogroup comparison ($p = 0.5182$); however, the sample size for each subgroup was comparatively small.

Proportions of the E, Q, R, and remaining haplogroups are shown in Figure 3. All have a substantial European (haplogroup R) composition. The probability distributions of the primarily African haplogroup E were not significantly different between the three populations. Overall, haplogroup E represents ~19% of the total haplogroup composition for all populations. The ABI population has the highest proportion of haplogroup R at 55.24% ($\pm 4.85\%$). The Native Alaskan population has the lowest proportion of haplogroup R at 24.55% ($\pm 2.43\%$). Overall, haplogroup

R and its subclades represent ~31% of the total haplogroup composition for all populations. The Native Alaskan population shows the highest proportion of Native American haplogroup Q at 43.08% ($\pm 2.34\%$). The Sioux population shows the lowest proportion of haplogroup Q at 6.41% ($\pm 1.96\%$). Overall, haplogroup Q makes up ~31% of the combined haplogroup composition for all populations. The Sioux population is comprised mostly of individuals assigned to haplogroups other than E, Q, and R with a proportion of 39.74% ($\pm 3.92\%$). All other haplogroups account for ~20% of the total haplogroup composition for all populations. In aggregate, these suggest that in all three datasets of Native Americans analyzed here, a substantial proportion of their Y-chromosomes are of non-Native American patrilineal ancestry.

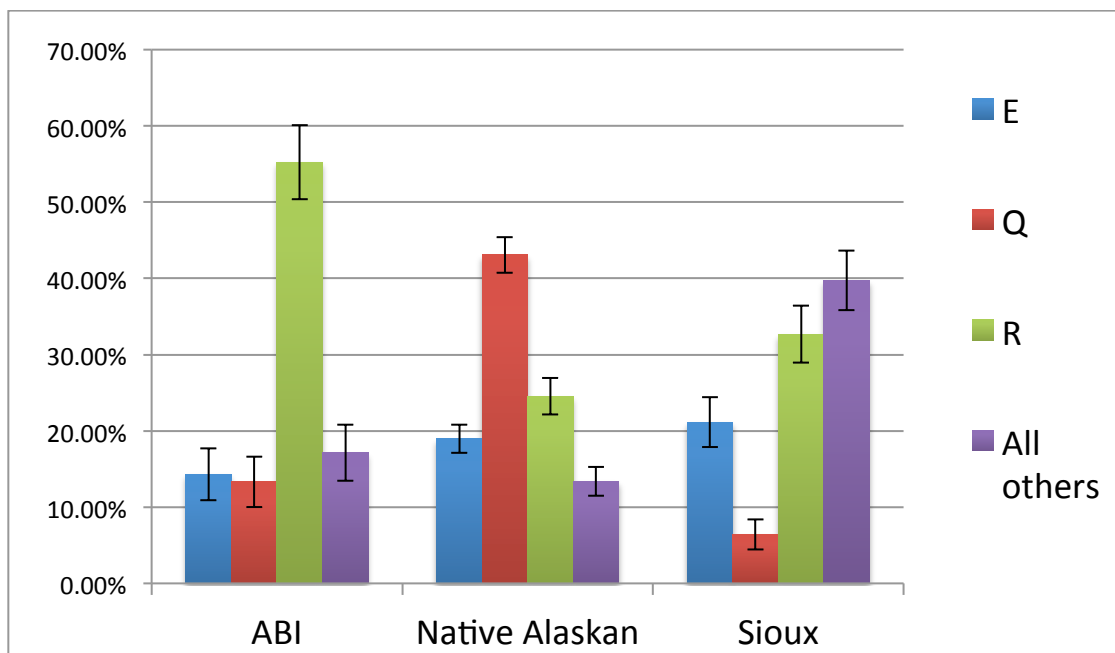


Figure 3. Haplogroup proportions of African (E), Native American (Q), European (R), and all others for ABI, Native Alaskan, and Sioux populations.

3.2 STRUCTURE

Figure 4 shows a triangle plot of individual membership coefficients in each cluster. Distances from each dot to one of three corners indicate the relative proportion of that cluster. Those individuals in corners were assigned completely to that cluster whereas admixed individuals fall in between corners. Based on the observed distribution of individuals, Cluster 2 represents the Native American haplogroup Q. The majority of individuals assigned to this cluster belong to the Native Alaskan populations, which is in accordance with its high proportion of haplogroup Q and the comparatively low proportion of Q in the ABI and Sioux populations (Figure 3). Cluster 1 most likely represents haplogroup R and African and all other haplogroups are therefore grouped into Cluster 3, given the high assigned proportion of Sioux individuals. The majority of Sioux and a substantial portion of ABI individuals were assigned to this cluster.

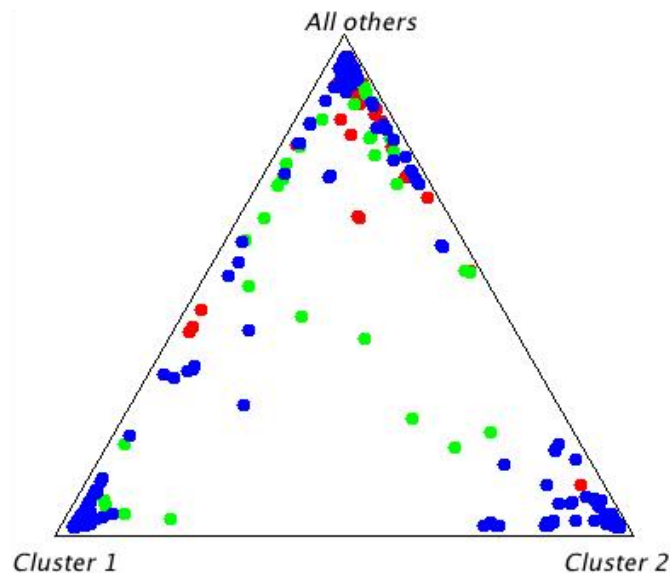


Figure 4. Triangle plot of individual membership coefficients. Individuals are represented by dots colored according to their population of origin: green = ABI; red = Sioux; blue = Native Alaskan. Cluster 1 = European; Cluster 2 = Native American; All others (Cluster 3) = African and all others.

Overall, 148/156 Sioux individuals (94.87%) were clearly assigned to one cluster with a probability of greater than 80%. The remaining individuals were jointly assigned to two or three clusters denoting individual admixture. Of the 148 individuals, 99 were assigned to Cluster 3 (66.89%). A total of 89/105 (84.76%) ABI individuals were assigned completely to either Cluster 1 or Cluster 3 indicating this population has the highest percentage of admixed individuals based on STRUCTURE results. Cluster 1 has a total membership of 60.67% and Cluster 3 has a total membership of 39.33%. Approximately 92.41% (414/448) of Native Alaskans were assigned unambiguously to one of the three clusters. Cluster 2 has the highest proportion of membership in Native Alaskans at 46.62%. Clusters 1 and 3 represent 25.36% and 28.02% of the Native Alaskan population, respectively. Membership proportions in each cluster averaged over all individuals for the three populations are shown in Table 5. Individual memberships generally agree with the averaged population memberships.

Population	Cluster 1	Cluster 2	Cluster 3
Sioux	0.323 ± 0.037	0.035 ± 0.015	0.642 ± 0.038
ABI	0.538 ± 0.049	0.056 ± 0.022	0.405 ± 0.048
Native Alaskan	0.258 ± 0.021	0.454 ± 0.024	0.289 ± 0.021

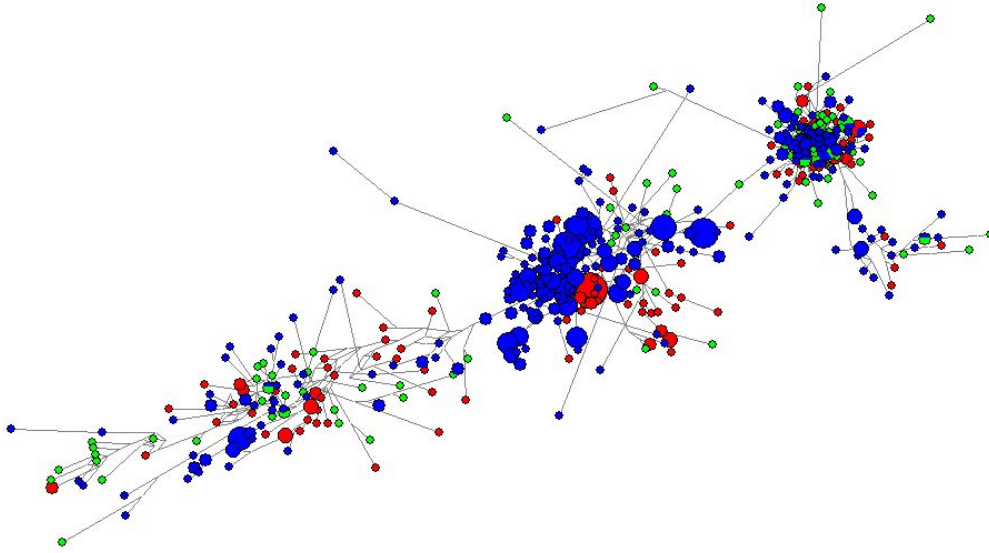
Table 5. Average proportion of membership of each population in each of the 3 clusters.

3.3 NETWORK

The phylogenetic relationship between individual haplotypes for ABI, Native Alaskan, and Sioux populations based on population of origin and haplogroup designation is shown in Figure 5. Most individuals from the ABI, Sioux, and Native Alaskan populations were placed in

three discrete clusters (Figure 5A). ABI and Sioux individuals appear to be more evenly distributed between the three clusters, whereas Native Alaskan individuals are more highly clustered into one group. Organizing individuals according to their assigned haplogroup yields a network with pronounced clustering of the R and Q haplogroups (Figure 5B). Haplogroup E and all others show a lower degree of clustering and are more distributed across the network. The Native Alaskan population shows a significant level of intra-population haplotype sharing. This is also seen to a lesser extent in the Sioux (Table 4). Inter-population haplotype sharing is limited to the ABI and Alaskan populations, visualized in Figure 5A as mixed red and blue circles (see also Table 4). The Sioux population sampled is composed of individuals whose haplotypes are exclusive to that population, though several are shared among individuals of different haplogroups.

A.



B.

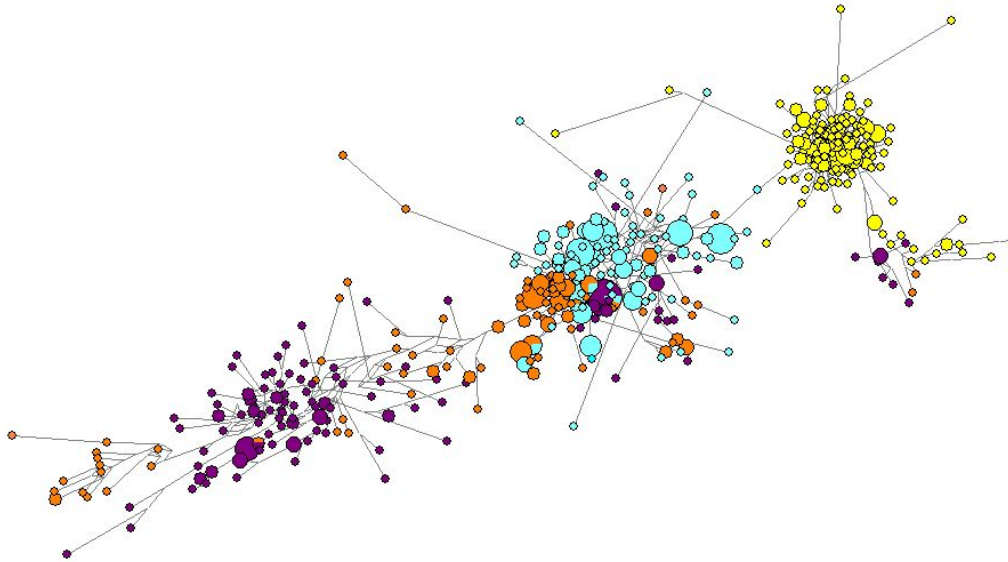


Figure 5. (*A*) Median-joining network using 12 Y-STR loci relating haplotypes of individuals from ABI (green), Native Alaskan (blue), and Sioux (red) populations. (*B*) Median-joining network using 12 Y-STR loci relating haplotypes of individuals grouped by haplogroup designation where Q = light blue, R = yellow, E = orange, and all others = purple. Circles represent haplotypes with an area proportional to frequency.

CHAPTER FOUR: DISCUSSION

This study characterized the variation of the non-recombining portion of the Y-chromosome for three sample populations: Sioux, Native Alaskan, and unclassified Native Americans. The ABI population showed the highest degree of Y-STR haplotype diversity as evidenced by the low incidence of haplotype sharing. The proportion of Y-chromosomes with Native American, European, and other ancestries is significantly different both among the three populations and among the three Native Alaskan subpopulations, confirming the presence of inter-ethnic admixture. Generalized admixture proportions as inferred from analysis of Y haplogroup distributions reveals the presence of European patrilineal ancestry, the largest of which is seen in the ABI population. However, Native American and all other haplogroups make up the majority of Native Alaskan and Sioux individuals in these samples, respectively.

STRUCTURE analysis further confirmed the presence of admixture and some degree of substructuring in all three population samples. Individuals were not assigned completely to one cluster and one cluster only; instead, membership was usually divided between two and sometimes three clusters – a clear indication of individual admixture. The same can be said of the population membership proportions averaged across individuals. No one population was unambiguously assigned to a single cluster. Each population had a majority cluster assignment; however, a substantial portion of the populations was also assigned to another cluster. The Sioux population showed the highest level of African plus all other haplogroups admixture based on *STRUCTURE* results, which is in agreement with the calculated

haplogroup proportions. Native Alaskans showed the highest incidence of Native American affiliation supporting the conclusions drawn from haplogroup distributions. Populations were divided into three clusters corresponding to three major haplogroup groupings: African and all others, European, and Native American.

Median-joining network analysis indicates that the ABI, Sioux, and Native Alaskan sample populations are composed of three relatively genetically distinct groups.

In summation, analyses done in this study indicate that Native Americans of North America do not appear to have a homogeneous patrilineal ancestry. A substantial proportion of their Y-chromosomes is of non-Native American ancestry, and that proportion varies by tribal affiliation of Native Americans. Consequently, rarity of any specific Y-STR haplotype cannot be accurately evaluated only by examining Y-STR databases of Native American samples. This is so, because the specific target haplotype, by itself, may not be of Native American ancestry, and this study showed that the chance of this is not trivial, even when geographic populations of the group are considered for sampling (such as the Sioux sample studied here). For samples with no tribal affiliation described (such as the ABI dataset analyzed here), the chance of encountering Y-STR haplotypes of non-Native American ancestry is even higher.

To account for substructure present in these populations in statistical calculations of the rarity of any given Y-STR haplotype, a theta (θ) correction should be used. For Y-haplotype data, the value of θ depends not only on the number of loci the haplotype contains, but the specific loci of which the haplotype is composed as well. As the number of loci increases, θ decreases (32). Therefore, an appropriate value of θ should be employed based on the loci typed for the three Native American populations.

LIST OF REFERENCES

1. 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 3;64(3):817-31.
2. Fagundes NJR, Kanitz R, Eckert R, Valls ACS, Bogo MR, Salzano FM, et al. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet.* 2008;82(3):583-92.
3. Ray N, Wegmann D, Fagundes NJR, Wang S, Ruiz-Linares A, Excoffier L. A statistical evaluation of models for the initial settlement of the American continent emphasizes the importance of gene flow with Asia. *Mol Biol Evol.* 2010;27(2):337-345.
4. Bortolini M-, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, et al. Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet.* 2003;73(3):524-39.
5. Corach D, Lao O, Bobillo C, Van Der Gaag K, Zuniga S, Vermeulen M, et al. Inferring continental ancestry of Argentineans from autosomal, Y-chromosomal and mitochondrial DNA. *Ann Hum Genet.* 2010;74(1):65-76.
6. Dulik M, Zhadanov S, Osipova L, Askapuli A, Gau L, Gokcumen O, et al. Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and indigenous Altaians. *Am J Hum Genet.* 2012 2/10;90(2):229-46.
7. Johnson M, Smith J. *Tribes of the Sioux Nation.* Osprey Publishing: Oxford 2000.
8. Smith SJ. *The Y-haplotypes of the South Dakota Native American Sioux.*[dissertation]. Orlando, Florida: University of Central Florida; 2003.
9. Davis C, Ge J, Chidambaram A, King J, Turnbough M, Collins M, et al. Y-STR loci diversity in Native Alaskan populations. *Int J Legal Med.* 2011;125(4):559-63.
10. <http://acrh-ahc.uaa.alaska.edu/ahc/clinical.html>

11. Butler JM. Recent developments in Y-short tandem repeat and Y-single nucleotide polymorphism analysis. *Forensic Sci Rev.* 2003;15(2).
12. Mulero JJ, Chang CW, Calandro LM, Green RL, Li Y, Johnson CL, et al. Development and validation of the AmpFℓSTR® Yfiler™ PCR Amplification kit: A male specific, single amplification 17 Y-STR multiplex system. *J Forensic Sci.* 2006;51(1):64-75.
13. <http://www.ncjrs.gov/App/publications/abstract.aspx?ID=233445>
14. Athey TW. Haplogroup prediction from Y-STR values using an allele-frequency approach. *Journal of Genetic Genealogy.* 2005;1:1-7.
15. <http://www.hprg.com/hapest5/>
16. Athey TW. Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. *Journal of Genetic Genealogy.* 2006;2(2):34-9.
17. Muzzio M, Ramallo V, Motti JMB, Santos MR, Camelo JSL, Bailliet G. Software for Y-haplogroup predictions: a word of caution. *Int J Legal Med.* 2011;125:143-147.
18. Blanco-Verea A, Jaime JC, Brión M, Carracedo A. Y-chromosome lineages in Native South American population. *Forensic Sci Int.: Genetics.* 2010 4;4(3):187-93.
19. Dupuy BM, Stenersen M, Lu TT, Olaisen B. Geographical heterogeneity of Y-chromosomal lineages in Norway. *Forensic Sci Int.* 2006 12/1;164(1):10-9.
20. Underhill PA, Kivisild T. Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annual Review of Genetics* 2007;41:539-564.
21. Jobling MA, Tyler-Smith C. The human Y chromosome: An evolutionary marker comes of age. *Nat Rev Gen.* 2003;4(8):598-612.
22. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. Genetic structure of human populations. *Science.* 2002;298(5602):2381-5.

23. Hammer MF, Chamberlain VF, Kearney VF, Stover D, Zhang G, Karafet T, et al. Population structure of Y chromosome SNP haplogroups in the United States and forensic implications for constructing Y chromosome STR databases. *Forensic Sci Int.* 2006 12/1;164(1):45-55.
24. Budowle B, Ge J, Aranda XG, Planz JV, Eisenberg AJ, Chakraborty R. Texas population substructure and its impact on estimating the rarity of Y STR haplotypes from DNA evidence. *J Forensic Sci.* 2009;54(5):1016-21.
25. <http://www6.appliedbiosystems.com/yfilerdatabase/>
26. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources.* 2010;10(3):564-7.
27. Roff DA and Bentzen P. The statistical analysis of mitochondrial DNA polymorphisms: Chi 2 and the problems of small samples. *Mol. Biol. Evol.* 1989;6(5):539-545
28. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155(2):945-59.
29. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics.* 2003;164(4):1567-87.
30. <http://www.fluxus-engineering.com/sharenet.htm>
31. Bandelt H, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 1999;16(1):37-48.
32. Ge J, Budowle B, Planz JV, Eisenberg AJ, Ballantyne J, Chakraborty R. US forensic Y-chromosome short tandem repeats database. *Legal Medicine.* 2010:289-295.