Forensic Science International: Genetics 22 (2016) 139–148

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

The mitochondrial landscape of African Americans: An examination of more than 2500 control region haplotypes from 22 U.S. locations

M. Scheible[a,b,**,1], R. Just[a,b,2], K. Sturk-Andreaggi[a,b,*], J. Saunier[a,b], W. Parson[c,d], T. Parsons[a,b,d,3], M. Coble[a,b,4], J. Irwin[a,b,2]

[a] Armed Forces DNA Identification Laboratory, 115 Purple Heart Dr., Dover AFB, DE 19902, United States
[b] American Registry of Pathology, 9210 Corporate Blvd. Suite 120, Rockville, MD 20850, United States
[c] Institute of Legal Medicine, Innsbruck Medical University, Müllerstrasse 44, Innsbruck, Austria
[d] Penn State Eberly College of Science, 517 Thomas Building, University Park, PA 16802, United States

ARTICLE INFO

Article history:
Received 30 August 2015
Received in revised form 14 December 2015
Accepted 6 January 2016
Available online 12 January 2016

Keywords:
Mitochondrial DNA
Control region
African American

ABSTRACT

The mitochondrial DNA (mtDNA) control region (16024-576) was Sanger-sequenced for a total of 2563 self-identified African Americans, using automated processing techniques and data review standards exceeding guidelines for forensic applications. Genetic diversity ranged from 0.9952 to 0.9998 in 22 population samples from 20 different states.

Haplogroups of African ancestry, found in 82.48% of individuals overall, were most concentrated in the Southeast U.S. and decreased to the north and west. West African and West Central African haplotypes were well-represented in the population samples, especially in the southern U.S. states, while East African haplogroups were observed in low-frequency clusters in a handful of locations across the country. East Asian, Native American, and West Eurasian admixture was present in 3.16%, 2.93%, and 11.43% of samples, respectively. While some geographic substructure was detected across the population samples as clines in admixture frequencies, 20 of the 22 population samples were found to be statistically indistinguishable by pairwise comparisons and AMOVA calculations. Datasets from Hawaii and Idaho, however, were clear outliers. Overall, these more than 2500 control region sequences represent the most comprehensive regional sampling of African American mtDNA diversity to date, and are suitable for use in a forensic mtDNA database. The population data are made available via EMPOP (www.empop.org) and GenBank.

1. Introduction

In forensics, significance is assigned to the results of a mitochondrial DNA (mtDNA) match comparison by estimating the frequency of the mtDNA haplotype within a relevant population. If the haplotypes of a known and unknown specimen are consistent, the observed frequency of the haplotype in a relevant population database provides an unbiased estimate of the likelihood that the specimens originated from the same maternal lineage. These frequency estimates and subsequent likelihood calculations thus depend upon the size of the reference population database. Moreover, it is imperative that the population structure in question has been examined to determine the extent of regional or location-specific variation, and whether databases from diverse populations samples can be appropriately pooled for the purpose of assessing the significance of a match.

The Trans-Atlantic Slave Trade, which took place during the 16th through 19th centuries, dramatically changed the genetic structure of the Americas. While the largest number of African captives disembarked in the Caribbean and the eastern coast of South America, an estimated 400,000 traveled to North America's mainland, nearly three quarters of whom originated from West Africa [1]. Just over 200,000 individuals are reported to have been taken to the Carolinas and Georgia, while 128,000 arrived in the Chesapeake region, and over 40,000 were taken to the northern

* Corresponding author. Fax: +1 302 346 8870.
** Corresponding author.
E-mail addresses: mkscheib@ncsu.edu (M. Scheible), kimberly.s.andreaggi.ctr@mail.mil (K. Sturk-Andreaggi).
[1] Present address: Forensic Sciences Institute and Department of Molecular Biomedical Sciences, North Carolina State University, 1060 William Moore Drive, Raleigh, NC 27607, United States.
[2] Present address: Federal Bureau of Investigation, 2501 Investigation Parkway, Quantico, VA 22135, United States.
[3] Present address: International Commission on Missing Persons, Alipašina 45A, 71000 Sarajevo, Bosnia and Herzegovina.
[4] Present address: National Institute of Standards and Technology, 100 Bureau Dr., M/S 8312, Gaithersburg, MD 20899, United States.

http://dx.doi.org/10.1016/j.fsigen.2016.01.002
1872-4973/

U.S. and the Gulf Coast [1]. Consistent with these historical records, genetic studies have suggested that the majority of Africans who were forced to migrate to the Americas originated from West and West Central Africa [2–8], though genetic influences from Southwest Africa and Southeast Africa have also been described [3–5].

Regional substructure exists in Africa [2,4,9,10], and the African region from which slaves were sought was often dependent on the crops that were common between African and American regions [10]. As a result, the substructure present in Africa was transferred to the Americas, with clusters of genetically-similar African Americans populating areas near the ports where they disembarked [10]. After the initial migration from various locations in Africa to specific ports in the United States, the later "Great Migration" (from approximately 1910–1930) when African Americans traveled north via particular routes (e.g., along the Mississippi River) would be expected to further affect the present day genetic population structure of African Americans [10]. Westward movement could have introduced additional geographic variation. Lastly, widespread admixture with individuals of European or Native American ancestry [10,11] has certainly contributed significantly to the mtDNA variability in self-described African Americans.

Previous genetic studies have in fact demonstrated differing European contributions to self-described African American populations in the U.S., with significant variation observed across not only the particular geographic regions sampled, but also the genetic markers typed. Analyses of autosomal markers have indicated wide variation in European influence (from 3.5% to 24% [3,12–17]), and Y chromosome studies have reported an even broader range of 5.6–46.9% European ancestry proportions across the country [8,12,13,16,18]. The European influence detected in maternal lineages has ranged from 0% to 14.9% [12,13,16,19–21]—a considerably smaller spread than has been revealed in analyses of patrilineal or diploid markers. Two mtDNA studies that sampled from specific cities and inferred the proportion of European ancestry by haplogroup-specific restriction fragment length polymorphisms found higher European mtDNA admixture in the northern half of the country ($8.0\% \pm 5.0\%$) and lower European influence in the South ($4.2\% \pm 3.0\%$) [12,13]; however, these examinations were limited to just eight states in the Midwest and Eastern U.S. Contrary to these variable European influences, all studies to date, across all markers examined, have demonstrated low East Asian and Native American contributions to the African American population (0–4.9%) [12–14,17,19,21], with the high end of the range detected in a sample from the western U.S [17].

In the context of forensic mtDNA population reference data, African American diversity has been examined using a variety of methods with varying degrees of resolution, including sequence-specific oligonucleotide typing [8], single nucleotide polymorphism typing [22], Sanger-based studies ranging from hypervariable region only sequencing to full mitochondrial genome typing [18–21,23], and massively parallel sequencing [24]. While some regional sampling was performed in a few of those examinations, either no geographic substructure was detected [18] or only minor geographic differentiation was identified in populations on opposite U.S. coasts [8].

To both address the question of potential geographic substructure within African American populations, as well as to improve the size and breadth of African American mtDNA reference datasets for forensic applications, we conducted a comprehensive study of mtDNA diversity among self-described African Americans. Complete mtDNA CR sequences were developed for 2563 individuals sampled from 22 U.S. locations, using automated laboratory protocols. Data review procedures were applied to produce forensic-quality data [25] that meet all current best practices prescribed by the Scientific Working Group on DNA Analysis Methods (SWGDAM; [26]) and the DNA Commission of the International Society for Forensic Genetics (ISFG; [27]). In this report we examine both standard genetic diversity indices and the extent of regional heterogeneity in the most comprehensive regional sampling of African American control region haplotypes to date.

2. Materials and methods

2.1. Samples

Anonymized samples from 22 locations across 20 different states (Table 1) were obtained from various contributors. Specimen

Table 1
Sample information.

Population	Population size	Sample size	Sample type	EMPOP accession	GenBank accession
Alabama	110	110	Whole blood	EMP00047	KP319480–KP319589
Alaska	100	100	Blood serum	EMP00444	KP319380–KP319479
Mesa, Arizona	100	100	Whole blood	EMP00511	KP319681–KP319780
Phoenix, Arizona	91	91	DNA extract	EMP00453	KP319590–KP319680
Northern California	46	46	DNA extract	EMP00456	KP319781–KP319826
Orange County, California	362	362	Blood stain	EMP00593	KP319827–KP320188
Colorado	123	123	DNA extract	EMP00408	KP320189–KP320311
Connecticut	108	58	DNA extract	EMP00462	KP320312–KP320419
		50	Blood serum	EMP00570	
Florida	93	93	DNA extract	EMP00465	KP320420–KP320512
Hawaii	100	100	Blood serum	EMP00467	KP320513 - KP320612
Idaho	100	100	Blood serum	EMP00471	KP320613–KP320712
Illinois	100	100	Blood serum	EMP00474	KP320713 –KP320812
Minnesota	185	185	DNA extract	EMP00402	KP320813–KP320997
Missouri	98	98	Whole blood	EMP00478	KP320998–KP321095
Nebraska	107	107	Blood stain	EMP00487	KP321202 –KP321308
New York	140	40	DNA extract	EMP00492	KP321309–KP321448
		100	Blood serum	EMP00573	
North Carolina	106	8	DNA extract	EMP00483	KP321096–KP321201
		98	Blood stain	EMP00483	
Ohio	92	92	DNA extract	EMP00496	KP321449–KP321540
South Dakota	57	57	Blood stain	EMP00499	KP321541–KP321597
Texas	99	99	Blood serum	EMP00526	KP321598–KP321696
Vermont	147	147	DNA extract	EMP00420	KP321697–KP321843
Washington	99	99	Blood serum	EMP00530	KP321844–KP321942