# Mitochondrial DNA diversity in the African American population

**Derek C. Johnson**, **Sadeep Shrestha**, **Howard W. Wiener**, **Robert Makowsky**, **Ashish Kurundkar**, **Craig M. Wilson**, and **Brahim Aissani**

School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA

## Abstract

Genetic polymorphism along mitochondrial DNA (mtDNA) defines population-specific signatures called mtDNA haplogroups. Estimation of mtDNA haplogroup distribution may be prone to errors, notably if the study sample is not drawn from a multicenter cohort. Here, we report on mtDNA diversity in a sample of African American individuals ($n = 343$) enrolled in a multicenter cohort. Sequencing of the hypervariable regions I and II of the D-loop control region showed that the most common mitochondrial variants are 73G, 146C, 150T, 152C, 189G, 16278T, and 16311C. In agreement with the published data, we observed 17 common mtDNA haplogroups: L0, L1, L1b, L1c, L2, L2a, L2b, L2c, L2e, L3, L3b, L3d, L3e, L3f, L3h, L3x, and L4. The most commonly observed haplogroup is L2a (19.8%), followed by L1b (10.2%). Overall, the observed mtDNA haplogroup distribution in our study is similar to those published for the African American and the African populations.

## Keywords

Admixture; African American; mtDNA haplogroup; polymorphism; REACH

## Introduction

Human mitochondrial DNA (mtDNA) is a valuable source of information for studying genetic diversity in human populations. mtDNA is inherited maternally and is characterized by mutation rates many times greater than those of most nuclear DNA (Brown et al., 1979). The most polymorphic sites of mtDNA are concentrated in the non-coding hypervariable regions I (HVI) and II (HVII) of the D-loop control region (Vigilant et al., 1989). HVI and HVII have been very useful in phylogenetic studies and forensic science. HVI and HVII haplotypes are differentiated in common groups called mtDNA haplogroups representing the major branching points on the human mitochondrial phylogenetic tree. Recent advances in

technologies allowed high throughput sequencing of mtDNA from diverse human lineages and has helped trace the matrilineal inheritance of modern humans back to human origins in Africa and the subsequent spread out to the Asian and European continents (Soares et al., 2009).

Few studies have described the variability of mtDNA in the African American population in United States (Diegoli et al., 2009; Ely et al., 2006; Salas et al., 2002). A comprehensive analysis of mtDNA in a sample of 1148 African Americans reported a distribution of mtDNA haplogroups and relevant single nucleotide polymorphisms (SNPs) similar to those observed in African populations (Allard et al., 2005). Thirteen of the 18 haplogroups previously observed in African populations were observed in the African American populations: L1a, L1b, L1c, L2a, L2b, L2c, L3b, L3d, L3e1, L3e2, L3e3, L3e4, and L3f. Other studies based on HVI sequences have found more than half of their study samples of African American descent to share common haplogroups with African ethnic groups, particularly from West and West Central Africa (Bandelt et al., 2001; Salas et al., 2004). Genotype data for more than 450 000 single-nucleotide polymorphisms (SNPs) typed in Africans residing in different geographic regions, as well as African Americans and European Americans participating in the Atherosclerotic Disease Vascular Function and Genetic Epidemiology (ADVANCE) study confirmed the West and West Central Africa origins for the majority of the African American participants (Zakharia et al., 2009).

The study populations sampled in these few investigations of mtDNA haplogroup diversity in the African American population (Diegoli et al., 2009; Salas et al., 2002) may not all be representative of the African American population. Studies examining mtDNA diversity based on samples of individuals recruited from a single city or region of residence may not be representative of the general population because of the heterogeneity of the African American population. In this study, we have studied mtDNA diversity in African American individuals enrolled in a multicenter cohort study, Reaching for Excellence in Adolescent Care and Health (REACH). The REACH cohort study recruited individuals from 15 clinical centers in 13 major metropolitan cities throughout the United States (Rogers et al., 1998). The African American populations in these cities is diverse and can be traced back to geographic regions of Africa and the Caribbean (Ely et al., 2006; Zakharia et al., 2009). Thus, sampling individuals from metropolitan cities would widen the representation of the actual diversity of these populations. Studying the distribution of mtDNA haplogroups in individuals from multicenter cohorts such as REACH allows a more comprehensive exploration of mtDNA variation and generalization of the data. Here, we report data on mtDNA diversity in a representative population of African American obtained from full HVI and HVII sequences in a study sample of 343 individuals.

## Methods

### Study population

Our study population consisted of 343 African American individuals enrolled in the REACH study. The REACH study enrolled HIV-positive and HIV-negative at-risk adolescents between the ages of 12 and 18. Enrolment into REACH took place between February 1996 and November 1999 at 15 clinical programs in 13 cities throughout the United States.

Recruitment sites included four sites in the North-East, three sites in the Mid-Atlantic, six sites in the South-East, one in the Mid-West, and one in the South-West. The parent study and this sub-study conformed to the procedures for written informed consent (parental permission was obtained wherever required) approved by institutional review boards at all sponsoring organizations and to human-experimentation guidelines set forth by the United States Department of Health and Human Services. The REACH study has been previously described in detail (Rogers et al., 1998).

### DNA sequencing

Due to the presence of multiple copies of mtDNA in each cell, only tiny amount of DNA was necessary to amplify HVI and HVII, and initial DNA sequencing data indicated that the integrity of mtDNA was not affected. The HVI and HVII regions span positions 16,024 to 16,482 and 21 to 413, respectively. We have used the following PCR primers to amplify separately these polymorphic regions; HVI forward: 5′-TTAACTCCACCATTAG CACC-3′; HVI reverse: 5′-CCTGAAGTAGGAACCAGATG-3′, HVII forward: 5′-GGTCTATCACCCTATTAACCAC-3′; HVII reverse: 5′-CTGTTAAAAGTGCATACCGCCA-3′. "Basic local alignment search tool" (BLAST, release number 2.2.27+) analysis against the human reference genome and assemblies was performed with these PCR primer sequences to rule out possible annealing to segments of the nuclear genome. The PCR conditions for amplifying 20 ng of total DNA were 35 cycles of denaturation at 94 °C for 1 min, annealing at 56 °C for 1 min and extension at 72 °C for 1 min followed by a final extension at 72 °C for 10 min (Morovvati et al., 2007). PCR master mix for a 20 μl reaction consisted of 2 μl Buffer 10×, 2μl dNTP (200 μM), 1 μl each of R and F PCR primers (20 μM each), and 1 unit of Taq polymerase in a reaction of 20 μl. For quality control, 2 negative controls were incorporated in each 96-well plate. Prior to sequencing, unincorporated dNTPs and DNA primers remaining in the PCR products were removed by means of exonuclease I and Shrimp Alkaline Phosphatase (ExoSAP, United States Biochemical) treatments. Randomly chosen PCR products from each plate were tested by agarose gel electrophoresis to confirm the expected amplicon size. Big Dye terminator cycle sequencing was used to sequence single DNA amplicons. All PCR products were sequenced bidirectionally.

### Phylogeny analysis

The quality of forward and reverse sequences was calculated in CodonCode Aligner 3.7.1.1 and sequences with scores 300 were discarded. Consensus sequences were assembled using default conditions and poor quality regions were re-sequenced at least once. Regions that were not resolved by the sequencing assay (runs of "n" at both ends of the sequences or around the stretch of "C" in the HVI sequence) were coded as missing. The trimmed sequences were aligned using Clustal W2 (Larkin et al., 2007). Sequence data for the two hypervariable regions were concatenated and analyzed using a parallel version RaxML 7.2.6 on a Linux Cluster. Only unique haplotypes were included in the reconstruction of the phylogenetic tree, as duplicate haplotypes do not affect tree structure or support, thus reducing computation burden. All analyses were run assuming an unequal transition/transversion ratio and an equal mutation rate across sites. The likelihood analysis consisted of 100 separate heuristic searches and 1000 bootstrap replicates were performed using the

GTRCAT model of evolution. All sequence designations are based on comparisons to the revised Cambridge reference sequence (rCRS; Andrews et al., 1999).

## Haplogroup inference

Haplogroup assignments were determined through MitoTool (Ruiz-Pesini et al., 2007), a web-based bioinformatic platform designed for analyzing mtDNA in batch mode (Fan & Yao, 2011). MitoTool aligns HVI and HVII sequences with rCRS using Clustal W and exports the variants of each sequence in batch mode. Based on haplogroup-specific variation motifs, the haplogroup of each mtDNA is determined by optimal exact matching and fuzzy or near matching. We confirmed the haplogroup inferred in MitoTool by independent testing in the mthap program (Lick, 2013). Finally, we have combined classification data from MitoTool and mthap and our reconstructed phylogenetic tree (not shown) to cross-validate the inferred haplogroups and resolve discrepant results. Specifically, phylogenetically defined branches or haplogroups were used to assign a likely haplogroup to individuals for whom classification by MitoTool and mthap have failed or produced ambiguous haplogroups (more than one possible haplogroup per individual).

## Population composition

Self-report of population of origin and ethnicity was checked by discriminant analysis of a validated set of 150 ancestry informative markers (AIMs) typed in a previous REACH study (Shrestha et al., 2011). Individuals deemed to be of Hispanic-African American descent (Individuals clustering with the majority of self-reported Hispanic-African American descent) and individuals of admixed population descent other than non-Hispanic African American (individuals with two-component PCA coordinates falling outside the data clusters defining the above ethnic groups and the Caucasian group) were excluded. Furthermore, discriminant and PCA analyses were used to assess the correlation between mtDNA haplogroups and AIMs in their potential to differentiate population subgroups.

DNA variants were compared against two well established mtDNA databases; the Uppsala University "Human Mitochondrial Genome Database" (Ingman et al., 2000) and Yonsei University's mtDNA manager database (Lee et al., 2008). The Uppsala University "Human Mitochondrial Genome Database" is a comprehensive database of complete human mitochondrial genomes accumulated since year 2000. Recent expansions in the capabilities of Uppsala University's mtDB include a haplotype search function and the ability to identify and download sequences carrying particular variants (Ingman et al., 2000). Yonsei University's mtDNA database is a web-based application, which allows users to manage and analyze mtDNA information.

# Results

Principal component analysis based on a validated set of 150 nuclear AIMs typed in the predominantly African American REACH cohort study allowed to identify and exclude study participants of admixed population descent other than the African and non-Hispanic European descents (Figure 1). Individuals clustering with the majority of self-reported Hispanic-African Americans (those with two-component PCA coordinates falling between

the African American and Caucasian clusters) and individuals not belonging to any of these 3 ethnic groups (upmost right cluster in Figure 1) were excluded from further analyses. Inconsistencies between self-report of population of origin and PCA data clustering were observed for several individuals allowing some of them to be re-classified and either incorporated or excluded from the study. The re-classification defined a total set of 343 non-Hispanic African Americans available for this study. Furthermore, we compared the ability of mtDNA haplogroups and AIMs to differentiate population subgroups in a study sample that included all participants for whom both nuclear and mitochondrial typing data were available ($n = 352$), regardless of their populations of descent. The results showed a good concordance between the uniparental mtDNA haplogroups and biparental AIMs in assessing the composition of the study population. As can be seen with the colored data points, mtDNA haplogroups further split continental populations defined by nuclear AIMs in different subgroups (supplementary Figures S1A and S1B). The split is less evident for the European group because of the greater homogeneity of these populations compared to the African populations and also because Europeans are less represented in REACH and even less in the present sub-study than African Americans.

The differences in the distribution of mtDNA haplogroups between HIV+ ($n = 273$) and HIV− ($n = 70$) groups were not statistically significant ($p = 0.23$), allowing the pooling of these sub-groups. The distribution of haplogroups was similar to that found in the Africa's western cost (Salas et al., 2005), with predominance of haplogroups L0, L1b, L1c, L2a, L3b, L3d, and L3e. L2 (29.4%) and L3 (32.7%) were the most common macrohaplogroups (Figure 2).

## Distribution of L0, L1, L2, L3, and L4 macrohaplogroups

The L0 macrohaplogroup comprised 4.1% ($n = 14$) of the total macrohaplogroups. The L1 macrohaplogroup comprised 19.0% ($n = 65$) of the total macrohaplogroups, with the majority of its constituent haplogroups (53.0%) represented by L1b ($n = 35$) while the rest (47.0%) belonged to L1c ($n = 30$). The L2 macrohaplogroup (29.4%; $n = 101$) was predominantly (67.3%) represented by the L2a haplogroup ($n = 68$). Over 14% ($n = 15$) of the individuals in the L2 macrohaplogroup were classified as branching L2. The L3 macrohaplogroup comprised 32.7% ($n = 112$) of the total macrohaplogroups, making it the largest portion in our study. Individuals belonging to L3 were classified into 6 different haplogroups: L3b, L3d, L3e, L3f, L3h and L3x, with the L3d and L3e haplogroups representing 59.0% of this macrohaplogroup. Twenty-three individuals belonged to the undifferentiated L3 haplogroup (branching haplogroup). Two individuals were categorized as belonging to the L4 macrohaplogroup (Figure 2).

## Distribution of non-L haplogroups

A fraction (14.3%; $n = 49$) of our study population carried mtDNA haplogroups other than the African-specific L haplogroups. These non-L haplogroups consisted of R0 ($n = 24$; 6.9%), R ($n = 5$; 1.4%), M ($n = 8$; 2.3%), U ($n = 8$; 2.3%), and N ($n = 4$; 1.2%). The percentage of European and South Asian haplogroups in our population is comparable to other studies of African Americans (Allard et al., 2005; Figure 2).

## Single-nucleotide polymorphism by mitochondrial DNA haplogroup

SNP found in the major African American haplogroups are listed in Tables 1 and 2. Several nucleotide changes observed in our study were similar to those reported earlier for African and African American populations (Brehm et al., 2002; Chen et al., 2000; Pereira et al., 2001; Rando et al., 1998; Salas et al., 2002; Torroni et al., 2001). For all L-type mtDNA haplogroups, SNPs reported in this study have all been previously documented in the "Human Mitochondrial Genome Database" of Uppsala University (Ingman & Gyllensten, 2006) or in the mtDNA manager database of Yonsei University (Lee et al., 2008).

### L1b and L1c haplogroups

Haplogroup L1b was defined by 14 variant positions while the L1c haplogroup was defined by 18 variant positions (Tables 1 and 2). Our study found the following variant sites common to L1b and L1c: 73G, 152C, 182T, 195C, 247A, 16187T, 16293G, and 16311C.

Haplogroup L1b DNA variants at positions 182, 185, 195, 247, 16126, 16187, 16,278, and 16,311, and haplogroup L1c variants 152, 182, 189C, 247, 16,129, 16,187, and 16,311 have been reported in the mtDNA manager database of Yonsei University (Lee et al., 2008). DNA variants at positions 73, 16,187, 16,234, 16,265, 16,360 for L1b and those for L1c at positions 73, 198, 317, 16,192, 16,234, 16,360 were not listed in the mtDNA manager database of Yonsei Universit but were listed in the Human Mitochondrial Genome Database of Uppsala University (Ingman & Gyllensten, 2006).

### L2a, L2b, L2c, and L2e haplogroups

The L2a haplogroup was defined by 10 variant positions while the L2b haplogroup was defined by 14 variants positions (Tables 1 and 2). DNA variants positions common to L2a and L2b included: 73G, 146C, 152C, 195C, 198T, 16278T, and 16390A.

The L2a DNA variants at positions 146, 152, 195, 16278 and 16294, and the L2b DNA variants at positions 146, 150, 152, 182, 195, 198, 204, 16,129, 16,213, and 16,278 observed in our study were reported in the Yonsei University database. However, variants at positions 73, 16,192, 16,309 and 16,390 for L2a and 73, 16,355 and 16,390 for L2b were not listed in the Yonsei database; these variants were found on the same haplogroups in the Uppsala University database.

The L2c and L2e haplogroups represented a minor fraction of the total mtDNA haplogroups in our study population (Figure 2). The L2c haplogroup (1.1%, $n = 4$) was defined by the following 9 DNA variants while the L2e haplogroup (0.9%, $n = 3$) was defined by 11 variants (Tables 1 and 2).

### L3b, L3d, L3e, L3f haplogroups

Both the L3b and L3d haplogroups were defined by seven variant positions. The L3e haplogroup was defined by 12 variant positions and the L3f haplogroup was defined by six variant positions. While the only variant position shared by all three L3 haplotype subgroups was 73G, several variant positions were shared by three of the four L3 haplotype subgroups. DNA variant positions 150T and 195C were found in the L3d, L3e and L3f haplogroups.

Variant positions 152C was shared by L3b, L3d, and L3e and variant positions 189G was shared by L3b, L3e, and L3f haplogroups (Tables 1 and 2).

L3b DNA variants at positions 16,124, 16,278, and 16,362; L3d DNA variants at positions 152 and 16,124; L3e DNA variants at positions 150, 189, 200, and 16,327, and L3f DNA variants at positions 189, 200, 16,129, and 16,209 were reported in the Yonsei University database. Variants at positions 73, 152, 195, and 16,311 in L3b, at 73, 150, 159, 195, 16,111 and 16,399 in L3d, at 73, 152, 195, 198, 16,168, 16,264, and 16,320 in L3e, and at 73, 150, and 195 in L3f were not listed in the Yonsei database but in the Uppsala database.

## Discussion

The distribution of haplogroups in our data set is similar to those reported in other African American populations (Ely et al., 2006; Zakharia et al., 2009). In agreement with the published data (Ely et al., 2006; Salas et al., 2004) our study found that L2a is the most common haplogroup among African Americans. Many of the Africans brought to the Americas during the 18th century came from West Africa through the Atlantic Slave Trade; this explains the high rates of L2a haplotypes in both African Americans and West Africans (Salas et al., 2002). Some haplogroups in the REACH cohort were found in different proportions compared to other published populations. For example, 11% of individuals sampled in the SWGDAM multicenter study were of haplogroup L1c and 8.1% of L3b (Allard et al., 2005). By comparison, 8.7% and 3.8% of the African Americans in the REACH cohort were of L1c and L3b haplogroups, respectively. A plausible explanation to these discrepant frequencies is the smaller size of our sample compared to that of SWGDAM. Alternatively, these discrepancies can be explained by methodological differences. Our study only sampled African Americans who self-reported as non-Hispanics and assessed the race membership on the basis of AIMs (nuclear markers), whereas SWGDAM only used self-reported race and ethnicity to categorize their study population. While the self-report of race/ethnicity usually matches well with genetically defined ancestries, further exclusion of some individuals on the basis of PCA data in our study may explain the observed difference with the SWDAM data. Indeed, we identified a fair number of L3b individuals among the subgroup of African Americans with Hispanic ethnicity. The other possible explanation is a different proportion of admixed populations other than the African American population between SWDAM and our study. A fraction (17.1%) of our study population was identified as being potentially of admixed populations of origin other than the African and European populations This observation is substantiated by the overrepresentation of C1 and A2, two haplogroups highly prevalent in Southeast Asia and South and Central Americas; respectively, among individuals of Hispanic ethnicity. Of potential application as an alternative method to control for population structure in association studies or to be used for other purposes in anthropological and Forensic sciences, we show that the combination of a small but informative selection of AIMs (Shrestha et al., 2011) and mitochondrial haplogroups provides additional information on population composition. As expected from the high mutation rate and heterogeneity characterizing mtDNA, mtDNA haplogroups further breakdown the continental populations defined by AIMs into different subgroups. Thus, for large and powered genome-wide association

studies of populations of African descent, mtDNA haplogroups and more generally mtDNA macro-haplogroups can be used as covariates to control for population structure.

While the majority of the haplogroups found in our study population were of the African heritage L groups, 14.3% of them had non-L type mtDNA haplogroups. This is higher but still comparable to the frequencies of non-L mtDNA haplogroups reported in the SWGDAM study, which found non-L haplogroups to represent 8% of their sample (Allard et al., 2005).

The sequencing of the mtDNA HVI and HVII regions has allowed us to identify combinations of DNA variants specific to the L0, L1, L2, and L3 macrohaplogroups, which have been reported in the Human Mitochondrial Genome Database of Uppsala University and the mtDNA manager database of Yonsei University. For instance, variants 64T and 16230G belong to the L0 haplotype. DNA variant 247A, in combination with 185T and 357G, are unique to L1b, and variants 317A and 16234C belong to L1c. Jointly with variant 146C, 16192T defined L2a, 16129A defined L2b and16264T defined L2C. While no unique combination of DNA variants was common to all L3 sub-haplogroups, DNA variant 159C belonged to L3d, 16051G to L3e, and 16209C to L3f.

There was no representation of the L0a haplogroups in the REACH cohort. This contrasts with the mtDNA data from other multicenter cohorts of African American populations. In the SWGDAM study (Allard et al., 2005), 3.6% of the participants carried the haplogroup L0a (defined as L1a in the older nomenclature (Mishmar et al., 2003)). It is possible that some positions defining this haplogroup were not resolved in our sequencing assay or were not covered by our selected primers. Compared to our study, which sequenced the HVI and HVII regions 16024–16482 and 21–413, respectively, the SWGDAM study sequenced the entire control region. With the exception of variants 64T and 16230G, many of the diagnostic positions for L0a are not unique to the L0 macrohaplogroup, which is expected given that L0 is the common ancestor to all haplogroups. Our study shared the L0 diagnostic positions 185, 247, 16129, 16148, 16168, 16172, and 16187 with the Yonsei study. However, it did not share diagnostic positions 93, 236, 16188, 16189, 16223, and 16230. Other diagnostic positions for L0, such as variants 152C, 185A, 189G, 16311T, and 16320T were also found in the L2 and L3 haplogroups. Thus, it is possible that our study did not cover positions that further differentiate L0 and L0a individuals. The majority of L0a haplogroups in the Americas are found in South America where their lineage can be traced back to Mozambique (Pereira et al., 2001). Given that the REACH cohort samples are from 13 different US cities, it is unlikely that L0a haplogroup are confined to a particular geographic section of the US, which was not sampled by REACH. We have relied on both the self-report of race and principal component analysis to identify individuals of potentially admixed population of descent and to limit our study population to non-Hispanic African Americans.

As outlined above, the discrepancies between our study and SWDAM for L1b and L3b mtDNA haplogroups are likely attributable to differences in the sampling design (sampling from single or multicenter studies, inclusion or exclusion of individuals of admixed populations of origin other than strictly African and European or varying proportions of composite admixtures).

## Conclusion

Overall, the mtDNA diversity observed in our study is concordant with those published for the African American and the African populations. Specifically, the most commonly observed haplogroups found in our study were L2a (19.8%), followed by L1b (10.2%). The data reported here add to the increasing number of mtDNA sequences available for the African American populations and can be useful for forensic, anthropologic and biomedical investigations.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

## References

Allard MW, Polanskey D, Miller K, Wilson MR, Monson KL, Budowle B. Characterization of human control region sequences of the African American SWGDAM forensic mtDNA data set. Forensic Sci Int. 2005; 148:169–79. [PubMed: 15639612]

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 1999; 23:147. [PubMed: 10508508]

Bandelt HJ, Alves-Silva J, Guimaraes PE, Santos MS, Brehm A, Pereira L, Coppa A, et al. Phylogeography of the human mitochondrial haplogroup L3e: A snapshot of African prehistory and Atlantic slave trade. Ann Hum Genet. 2001; 65:549–63. [PubMed: 11851985]

Brehm A, Pereira L, Bandelt HJ, Prata MJ, Amorim A. Mitochondrial portrait of the Cabo Verde archipelago: The Senegambian outpost of Atlantic slave trade. Ann Hum Genet. 2002; 66:49–60. [PubMed: 12015000]

Brown WM, George M Jr, Wilson AC. Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci U S A. 1979; 76:1967–71. [PubMed: 109836]

Chen YS, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, Wallace DC. mtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations. Am J Hum Genet. 2000; 66:1362–83. [PubMed: 10739760]

Diegoli TM, Irwin JA, Just RS, Saunier JL, O'Callaghan JE, Parsons TJ. Mitochondrial control region sequences from an African American population sample. Forensic Sci Int Genet. 2009; 4:e45–52. [PubMed: 19948325]

Ely B, Wilson JL, Jackson F, Jackson BA. African-American mitochondrial DNAs often match mtDNAs found in multiple African ethnic groups. BMC Biol. 2006; 4:34. [PubMed: 17038170]

Fan L, Yao YG. MitoTool: A web server for the analysis and retrieval of human mitochondrial DNA sequence variations. Mitochondrion. 2011; 11:351–6. [PubMed: 20933105]

Ingman M, Gyllensten U. mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. Nucl Acids Res. 2006; 34:D749–51. [PubMed: 16381973]

Ingman M, Kaessmann H, Paabo S, Gyllensten U. Mitochondrial genome variation and the origin of modern humans. Nature. 2000; 408:708–13. [PubMed: 11130070]

Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, Mcwilliam H, Valentin F, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007; 23:2947–8. [PubMed: 17846036]

Lee HY, Song L, Ha E, Cho SB, Yang WI, Shin KJ. mtDNAmanager: A Web-based tool for the management and quality analysis of mitochondrial DNA control-region sequences. BMC Bioinformatics. 2008; 9:483. [PubMed: 19014619]

Lick, J. [last accessed 2013] Available from mthap. 2013. http://dna.jameslick.com/mthap/

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, et al. Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci U S A. 2003; 100:171–6. [PubMed: 12509511]

Morovvati S, Modarresi M, Habibi G, Kiarudi Y, Karami A, Peyvandi AA. Sequence analysis of mitochondrial DNA hypervariable regions: An approach to personal identification. Arch Med Res. 2007; 38:345–9. [PubMed: 17350487]

Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A. Prehistoric and historic traces in the mtDNA of Mozambique: Insights into the Bantu expansions and the slave trade. Ann Hum Genet. 2001; 65:439–58. [PubMed: 11806853]

Rando JC, Pinto F, Gonzalez AM, Hernandez M, Larruga JM, Cabrera VM, Bandelt HJ. Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. Ann Hum Genet. 1998; 62:531–50. [PubMed: 10363131]

Rogers AS, Futterman DK, Moscicki AB, Wilson CM, Ellenberg J, Vermund SH. The REACH Project of the Adolescent Medicine HIV/AIDS Research Network: Design, methods, and selected characteristics of participants. J Adolesc Health. 1998; 22:300–11. [PubMed: 9561462]

Ruiz-Pesini E, Lott MT, Procaccio V, Poole JC, Brandon MC, Mishmar D, Yi C, et al. An enhanced MITOMAP with a global mtDNA mutational phylogeny. Nucleic Acids Res. 2007; 35:D823–8. [PubMed: 17178747]

Salas A, Carracedo A, Richards M, Macaulay V. Charting the ancestry of African Americans. Am J Hum Genet. 2005; 77:676–80. [PubMed: 16175514]

Salas A, Richards M, De La Fe T, Lareu MV, Sobrino B, Sanchez-Diz P, Macaulay V, Carracedo A. The making of the African mtDNA landscape. Am J Hum Genet. 2002; 71:1082–111. [PubMed: 12395296]

Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A. The African diaspora: Mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet. 2004; 74:454–65. [PubMed: 14872407]

Shrestha S, Wiener HW, Olson AK, Edberg JC, Bowles NE, Patel H, et al. Functional FCGR2B gene variants influence intravenous immunoglobulin response in patients with Kawasaki disease. J Allergy Clin Immunol. 2011; 128:677–80. [PubMed: 21601260]

Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A, Salas A, Oppenheimer S, et al. Correcting for purifying selection: An improved human mitochondrial molecular clock. Am J Hum Genet. 2009; 84:740–59. [PubMed: 19500773]

Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, et al. Do the four clades of the mtDNA haplogroup L2 evolve at different rates? Am J Hum Genet. 2001; 69:1348–56. [PubMed: 11595973]

Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC. Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci U S A. 1989; 86:9350–4. [PubMed: 2594772]

Zakharia F, Basu A, Absher D, Assimes TL, Go AS, Hlatky MA, Iribarren C, et al. Characterizing the admixed African ancestry of African Americans. Genome Biol. 2009; 10:R141.1–R141.11. [PubMed: 20025784]
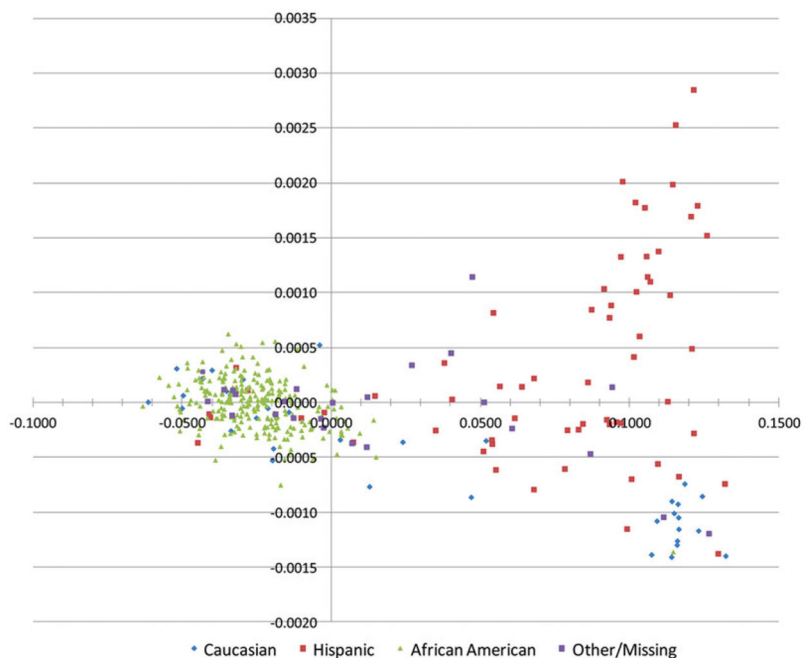
**Figure 1.**
Principal Component Analysis (PCA) and self-report of race in the REACH cohort. Two principal component-defined data clustering obtained from genotype data available for 150 Ancestry Informative Markers (AIMs). Self-report of race and ethnicity is indicated by colored symbols. Green triangles (African Americans); red squares (Hispanic ethnicity); purple squares (Other/missing) and blue diamonds (Caucasians). Inconsistencies between self-report of race and ethnicity and PCA clusters can easily be identified on the 2D PCA plot.
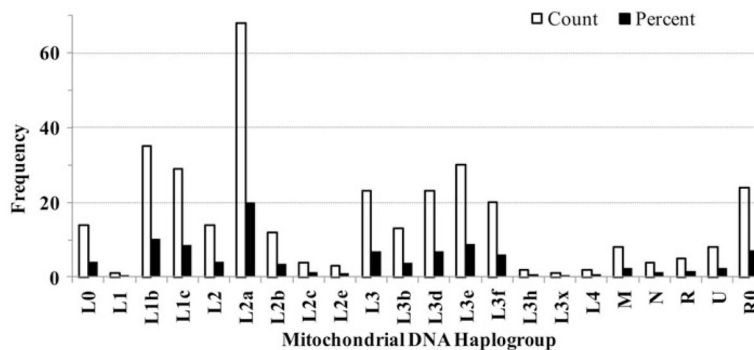
**Figure 2.**
Distribution of mtDNA haplogroups in the non-Hispanic African American population of the REACH cohort. Histograms showing the composition of mitochondrial DNA (mtDNA) haplogroups inferred from nucleotide sequences of the hypervariable regions I and II of the D-loop control region. mtDNA haplogroups were inferred via MitoTool and mthap, and discrepancies were resolved through the reconstructed phylogenetic tree.

**Table 1**

List of the DNA variant sites on HVII identifying major African American haplogroups in the REACH cohort.

| Haplogroup (%)[a] | 64T | 73G | 146C | 150T | 152C | 159C | 182T | 185A | 189G | 195C | 198T | 200G | 204C | 247A | 317A | 357G |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rCRS | C | A | T | C | T | T | C | G | A | T | C | A | T | G | C | A |
| L0 (4.1) | • | | | | | | | | | | | | | | | |
| L1b (10.2) | | • | | | • | | • | • | • | • | | • | | • | | • |
| L1c (8.5) | | • | | | • | | • | • | • | • | • | | | • | • | |
| L2 (4.4) | | • | • | | • | | • | | • | • | • | | | | | |
| L2a (19.8) | | • | • | • | • | | | | | • | • | | | | | |
| L2b (3.5) | | • | • | • | • | | • | | | • | • | | • | | | |
| L2c (1.1) | | • | • | • | • | | • | | | • | | | | | | |
| L2e (0.9) | | • | • | | • | | • | • | • | | | | | | | |
| L3 (6.7) | | • | • | • | | | • | | • | • | • | • | | | | |
| L3b (3.8) | | • | | | • | • | | • | • | | | | | | | |
| L3d (6.7) | | • | | • | • | | | | | • | | | | | | |
| L3e (8.8) | | • | | • | | | | | • | • | • | • | | | | |
| L3f (5.8) | | • | | • | | | | | • | • | | • | | | | |

[a]Percentage of individual haplogroup in the study population. Note that only the African-specific L lineages are shown. The dots represent DNA variant positions found within the haplogroups as shown on the first row of the table headings and in comparison to the rCRS reference alleles.

**Table 2**

List of the DNA variant sites on HVI identifying major African American haplogroups in the REACH cohort.

| Haplogroup (%)[a] | 16051G | 16093C | 16111T | 16124C | 16126C | 16129A | 16145A | 16148T | 16168T | 16172C | 16187T | 16192T | 16209C | 16213A | 16234T | 16230G | 16264T | 16265C | 16270T | 16278T | 16286G | 16293G | 16294T | 16309G | 16311C | 16320T | 16327T | 16355T | 16360T | 16362C | 16390A | 16399G |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rCRS | A | T | C | T | T | G | G | T | C | T | C | C | T | G | C | A | C | A | C | C | C | A | C | A | T | C | C | C | C | T | G | A |
| L0 (4.1) |  |  |  |  |  | • | • | • | • | • | • |  |  |  |  | • |  |  |  | • |  |  |  |  |  |  |  |  |  | • |  |  |
| L1b (10.2) |  |  |  | • |  |  |  |  | • |  | • |  |  |  |  |  |  | • |  | • | • | • |  |  | • | • |  |  |  |  |  |  |
| L1c (8.5) |  |  |  |  |  | • |  |  | • |  | • |  |  |  | • |  |  | • |  |  | • | • |  |  | • |  |  |  | • |  |  |  |
| L2 (4.4) |  |  |  |  |  |  |  |  |  |  |  | • |  |  |  |  | • |  |  | • |  |  | • | • |  |  |  |  |  |  | • |  |
| L2a (19.8) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | • |  |  |  | • |  |  | • | • |  |  |  |  |  |  | • |  |
| L2b (3.5) | • |  |  |  |  | • |  |  |  |  |  |  |  | • |  |  |  |  |  | • |  |  |  |  |  |  |  | • |  | • |  |  |
| L2c (1.1) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | • |  |  |  | • |  |  |  |  |  |  |  |  |  |  |  |  |
| L2e (0.9) |  |  | • |  |  |  | • |  |  |  |  |  |  |  |  |  |  |  |  | • |  |  |  |  |  | • |  | • |  |  | • |  |
| L3 (6.7) |  | • |  |  |  |  |  |  |  |  |  |  |  |  |  |  | • |  | • | • |  |  |  |  | • | • |  |  |  | • | • | • |
| L3b (3.8) |  |  | • | • |  |  |  |  |  |  |  |  |  |  |  |  |  |  | • |  |  |  |  |  | • |  |  |  |  | • | • |  |
| L3d (6.7) |  |  |  | • |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | • |
| L3e (8.8) | • |  |  |  |  |  |  |  |  | • |  |  | • |  |  |  | • |  | • |  |  |  |  |  | • |  | • |  |  |  |  |  |
| L3f (5.8) |  |  |  |  | • | • |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

[a] Percentage of individual haplogroups in the study population. Note that only the African-specific L lineages are shown. The dots represent DNA variant positions found within the haplogroups as shown on the first row of the table headings and in comparison to the rCRS reference alleles.