RECONSTRUCTING AFRICA'S EVOLUTIONARY HISTORIES: DNA COLLECTION, CODING, ANALYSIS, and INTERPRETATION

by

Taiye Winful

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Arts in
Anthropology

Charlotte

2018

Approved by:

_____
Dr. Jonathan Marks

_____
Dr. Lydia Light

_____
Dr. Fatimah Jackson

ABSTRACT

TAIYE WINFUL. Reconstructing Africa's Evolutionary Histories: DNA Collection, Coding, Analysis and Interpretation. (Under the direction of DR. JONATHAN MARKS)

Finding a way to successfully interpret African diversity is important in accurately reconstructing the evolutionary history of our species. By collecting high quality DNA, genealogical, and demographic data on a large cross section of African and African Americans a beginning bioculturally informed database was formed. By interpreting and contextualizing the data this database allows for further use in addressing health disparities among these communities. Africans and African Americans were the targeted population of this study. There was a wide range in both nationalities (26) and ethnicities (35) reported for the database. A majority of participants (56.2%) or about 260 participants self-reported their nationality as Black/African American. In total, there were 463 participants split up into four distinct databases based on sampling strategy that took ancestry into account.. 348 (75.2%) in AD (African Diasporic), 75 (16.2%) in AP Arabian Peninsula)(, 31 (6.7%) in CA (Continental African) and 9 (2.0%) in RA (Remote African). Gender, age, and geographical distributions were also examined. In conclusion, these databases have the potential to highlight the diversity of African and African American communities through contextualized bio histories, pave the way for precision medicine research and implementation, and lastly eliminate health disparities seen in these populations

ACKNOWLEDGMENTS

I would first like to thank my thesis advisor Dr. Jonathan Marks along with my other committee members Dr. Lydia Light and Dr. Fatimah Jackson and the Cobb Laboratory at Howard University.

I would also like to thank my friends and family who supported me throughout. This accomplishment would not have been possible without any of the people named. Thank you.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1: Introduction:

The broader purpose of this study is to develop a comprehensive bio culturally informed set of African DNA databases that reflect continental and diasporic African genomic diversity. The limited knowledge we have of African genetic/genomic variation is frequently ahistorical and not contextualized by African geospatial ecology and demography. It is important that DNA data be African-centered to accurately reconstruct the evolutionary history of our species because of the vast amount of genetic diversity within Africa. Currently we are unable to develop precision medicine and sophisticated genomic applications for peoples of recent African descent. With this in mind my research question is: How can African descended DNA collections be designed, collected, and analyzed effectively to reconstruct robust human evolutionary histories and address issues of contemporary health disparities? I hope to gain sight on the best sampling strategies to use for further investigation on addressing hypertension rates in these communities.

Aims and objectives of this study:

- Devise a set of sampling strategies for both Africans and African Americans to obtain a representative subset
- Collect adequate high quality DNA, genealogical, demographic, and clinical data on a broad cross-section of Africans and African Americans
- Code the derived data to produce a comprehensive biocultural database
- Prepare biological samples for analysis, and archive excess biological materials

- Obtain preliminary genomic data and conduct descriptive and inferential statistical analyses
- Interpret these analyses for evolutionary and clinical implications

Chapter 2: Literature Review:

Shortage of information about African and African American biological history:

Africa is a continent with a substantial amount of phenotypic, genetic, environmental and cultural diversity. Looking at modern human origin theories, according to the Out-of-Africa model, modern human species emerged from Africa and then dispersed throughout the rest of the world (Campbell and Tishkoff 2008, Tishkoff and Verrelli 2003). Modern humans originating from Africa makes African biological history essential in understanding modern human origins and genetic risk factors for disease (Campbell and Tishkoff 2008). Although the benefits of studying Africa are abundant, it is noticeably underrepresented in genetic studies (Reed and Tishkoff 2006, Campbell and Tishkoff 2008, Tishkoff *et al* 2009, Baharian *et al* 2015). Some genetic studies have used Africa in their work and have provided fruitful results (Schlebusch *et al* 2012). Genetic diversity has been shown to decline with distance from Africa (Tishkoff *et al* 2009). Much of what is known about African biological history is based on ethno-linguistic groups. Studying African populations offers insight into substantial genetic variation and adaptations to Africa's climate, diet, and exposure to infectious disease (Schlebusch *et al* 2012, Campbell and Tishkoff 2008). These variations lead to high levels of genetic and phenotypic variation in African populations (Reed and Tishkoff 2006).

The Role of the Transatlantic slave trade in African and African American Diversity:

Understanding African and African American migration could prove effective in understanding current biology (Zimmer 2016). In understanding African American

migration, there is a focus on 16th-19th century slaves due to high death rates and possible effects of natural selection due to slavery conditions (Wilson and Grim 1991). This time period was significant for two reasons: one, the transatlantic trade of African Americans and two, the institutionalization of slavery in the west. The transatlantic trade of millions of Africans to America resulted in many deaths; up to 10-15% of those forced into the trade, often recognized as the "middle passage", died on the ship and therefore did not make the full trip over to the US. Not only were mortality rates throughout the trip high but for those who did make it to America 10-30% of them died within 3 years of being enslaved. High mortality rates and low birth rates allowed for the continuation of the transatlantic slave trade. With this information, Wilson and Grim sought to answer the following: "How were the survivors of slavery different from those who died? Or, in other words, what traits were selected" (Wilson and Grim 1991: 125).

The main causes of death in slaves at this time included conditions like excessive vomiting, extreme sweating, extreme heat and diarrhea. These conditions often resulted in water and salt depletion. These same conditions existed on the plantations as well. At the time, standard treatment of these salt depleting diseases and conditions consisted of fluid and electrolyte replacement. Individuals with heightened ability to conserve fluids and electrolytes had a clear advantage over these conditions (Wilson and Grim 1991).

Variation among African and African Americans:

Researchers who want to study the relationships between genes and disease must be aware of the environments these individuals live in and come from. Detailed mapping of this variation will help show these gene-environment interactions (Zimmer 2016). It is

estimated that around 82.1% of current African Americans ancestors lived in Africa before the transatlantic trade travel while 16.7% of their ancestors resided in Europe and 1.2% in the Americas. Additionally, higher African ancestry is seen in southern states in comparison to northern and western ones (Baharian *et al* 2015). Genetic studies and gene environment interactions of African Americans are useful in generating informative data on functional variants, clarifying historical and genealogical mysteries, and showing basic biology.

Characterizing variation between African and African American populations may give a better understanding of disease susceptibility, health disparities, human evolutionary history and most importantly how genetic and environmental factors aid in producing different phenotypes (Campbell and Tishkoff 2008, Sirugo *et al* 2008). This may afford a significant example of human adaptability or developmental plasticity. The variation between the two populations comes from contextualizing their different life histories. Wilson and Grim use the previously mentioned transatlantic slave trade as a crucial role in the differences seen in these populations. More specifically, blood pressure observations allowed Wilson and Grim to identify fluctuations between African and West Hemisphere Black (African American) communities. Until then, it had not been widely taken into account that biological and environmental factors influence the genetic makeup of a population. Wilson and Grim use the life history of African Americans to explain the genetic and environmental factors that could play a role in African Americans predisposition to high blood pressure, in comparison with not only European Americans but African people as well (1991).

History of DNA collection in African Americans:

While today 45 million Americans self-identify as Black/African American, a majority of research linking most disease and genetics has, up until recently, been focused on people of European descent (Zimmer 2016). With the focus now shifting to those of other descents, African American collections are important. One comprehensive assessment looked at African American genomic diversity by looking at 3 genotype cohorts totaling 3,736 African Americans. These cohorts came from across the United States and provide an inclusive representative description of the United States as a whole (Baharian *et al* 2015). Through identity-by-descent analysis they were able to investigate nationwide population structure among African Americans and general relatedness. They used the cohort to understand the distribution of genetic variation and help aid in insight on human demography (Baharian *et al* 2015).

Understanding connections between history and DNA is fundamental for Africans and African Americans. As previously stated, genetic studies have historically focused on those of American and European history, but there has been a growing interest in larger scale studies (White 2016). A recently published study by researchers at McGill University looked at these larger scale studies to analyze the genetic history of African Americans. The Health Retirement Study and the Southern Community Concert Study are two of the largest databases of African American genetic data, both supported by the NIH. Looking at these studies they found that for most African Americans, a majority of their DNA can be traced back to Africa; interestingly they also found that African Americans in southern states have more African DNA than those living in other parts of the United States (Baharian *et al* 2015, White 2016).

Bioethical issues on collecting DNA information:

Participants' concerns about genetic research may affect their ability to enroll (Crider 2006). These issues often deal with consent, overall attitudes toward DNA collections and issues of ownership. The majority of labs that partake in DNA collection have no finite policies or agreements to address these issues properly (Godard *et al* 2003).

Consent:

Obtaining viable informed consent is a major issue in DNA collection (Deschenes *et al* 2001, McEwen 1996, Moutel *et al* 2001, Medical Research Council 2001). Informed consent is required when dealing with human biological materials (Godard *et al* 2003). It is important to pay attention to the cultural realities of different groups when dealing with ethical consent. In fact, informed consent is an essential aspect of contemporary health care practice because of its ability to be grounded in ethical principles of autonomy, confidentiality, and non-maleficence. Health challenges arise internationally or when dealing with different groups because beliefs about ethical requirements vary across cultures. One study investigated experiences of family caregivers of the elderly in Botswana. One aspect of the study focused on decision making when giving informed consent. They found that the use of Western decision-making models when dealing with the people of Botswana (or other cultures in general) lead to conflicting attitudes between Botswanan cultural values and western practice of health workers. Solutions to this issue include making sure informed consent is influenced by the realities of the social and cultural contexts in which clients are brought up in, and blending western and non-

western ideas of decision making (Shaibu 2007). Cozier's study focused on preferential DNA collection sample methods for black womens' health and he answered questions addressing approaches toward collection. One question looked at participation rate disparities with participants who received introductory and consent letters before or at the same time as their DNA collection kits. Surprisingly, they found that participants who read and signed their consent forms before they received their DNA collection kits were equally as likely to participate as those who received all at the same time (Cozier 2004).

Attitudes toward DNA collections:

There have been noticeable discrepancies in genetic research participation among different ethnic groups (Jenkins *et al* 2009). These discrepancies show that minority groups are less likely to participate in collections. Factors associated with participation are highly linked to race and ethnicity, however other secondary factors such as age, education and income play their respective roles. While minority groups are associated with decreased participation, white college educated participants are associated with increased participation (Crider 2006). Interestingly, African Americans, when compared to other ethnic groups are less likely to participate in genetic research (Crider 2006, Mcqullan *et al* 2006). Through focus group research, it was found that a main concern among these minority groups comes down to who is benefiting primarily from the research being done (Jenkins *et al* 2009, Crider 2006). These attitudes are often the result of unethical research done by the government as well as detrimental uses of the results (Jenkins *et al* 2009, Catz *et al* 2005).

Another major concern addresses how research samples can be linked with identifying sources. How these samples are linked to participants is detrimental in participation and assessing potential risks and benefits (Godard *et al* 2003, National Bioethical Advisory 1999). Samples are categorized into four categories: unidentified, unlinked, coded, and identified. The ASHG Ethics Committee recommended coding samples by a third party, such as storing the codes and keeping them inaccessible unless circumstances called for them to be disclosed (Godard *et al* 2003). To combat these preconceived attitudes, Jenkins *et al* 2009 looked for DNA collection methods that were convenient to mothers. They assembled focus groups with mothers who had gone through at home DNA collection kits for buccal swabs. The moderators asked the mothers questions concerning attitudes toward providing samples, factors that influenced participation, and preferred collection methods. They found that those who participated were more likely to participate for altruistic reasons and less likely to participate due to conspiracy theories, fear of government and unethical use (Jenkins *et al* 2009). Preferred collection methods were ones defined as being convenient and noninvasive; mouth rinsing methods ranked second (Jenkins *et al* 2009). Another study, conducted by Schwartz *et al* 2001, set out to find "whether attitudes about informed consent and willingness to participate in genetic research using stored DNA would depend upon the circumstances in which material is collected and characterized or disease under investigation (Schwartz *et al*, 337)." They found that a majority of participants (60-75%) believe consent should be required regardless of whether DNA was collected in a research or clinical setting. Further, they found participants were more likely to support

the need for consent in clinical settings more than in research settings (Schwartz *et al* 2001).

Ownership:

Issues of ownership emerge due to the possibility of commercial value in DNA samples (Godard *et al* 2003). The consensus about ownership of samples and accesses to biological information are determined by multi-party contracts, but not legislation (Martin and Kaye 1999, American College of Medical Genetics). Commonly, information given by a subject belongs to the research team, while the subjects themselves have no legal authority over their information (Godard *et al* 2003). Less popularly, some believe that subjects should own their samples even after it is collected and entered in a database. Another option suggests that potential subjects should decide whether they want to participate after they are informed about who will own the sample (Godard *et al* 2003). Concentrating on these factors is important if increased participation of minorities is to occur. Low rates of participation among minority groups, in this case African Americans, could limit researchers' ability to identify genes and gene-environment risk factor interactions (Jenkins *et al* 2009, Crider 2006). There has been a push to standardize ethical requirements and policies dealing with DNA collections (Godard *et al* 2003).

In addressing these issues it is also important to think about potential social, political, and ethical differences that these two populations could have. While these studies use the term "African-American" there is no distinction between African and

African American that I have acknowledged throughout this section. Understanding that these situations are unique to shared group issues will only add to better policy and agreement when dealing with bioethical issues concerning DNA collection.

Relationships between genetics and biocultural links to disease:

Slavery and systematic discrimination has led to an increase in social, economic, and health burdens in African American communities. These health burdens continue to worsen and create disparities when compared to other ethnic groups due to poverty, unequal access to care, and unequal medical representation for minority groups (Baharian *et al* 2015). To combat these disparities, cohorts in medical research that reflect and encompass the diversity in America are being formed at higher rates (Bustamante 2011, Burchard 2014).

Precision medicine involves prescribing individualized treatment that takes into account a patient's individual variability, including genetic, biomarker, phenotypic or psychosocial characteristics (Jameson and Longo 2015). The goal and benefit of precision medicine is the potential ability it has to enable better assessment of disease risk and disease mechanisms as well as the prediction of optimal therapy options. Ultimately, the main goal is to use precision medicine to produce a clinical database that can be used to benefit health and healthcare by anticipating, diagnosing, and treating disease. Another benefit of the database is the opportunity clients would have to access their own data and see how it is being used (Collins and Varmus 2015).

Addressing health disparities and disease susceptibility of hypertension in people of African descent:

Around 40% of African Americans, or people of African descent, are affected by hypertension (High Blood Pressure and African Americans 2017). Hypertension is defined as having abnormally high blood pressure (High blood pressure 2016, Cooper *et al* 1999, James 1991). Blood pressure is responsible for allowing the blood contracted by the heart to spread throughout the cardiovascular system (James and Brown 1997). As the heart contracts, it generates a pulse of blood. The average pulse of a person ranges around 100,000 systolic and diastolic pulse pressures in a day (James and Brown 1997, James and Pecker 1994). Systolic pressure is defined as being the maximum pressure that the pulse employs on the arterial wall of the heart while diastolic pressure is defined as minimum pressure of the heart resting just before its next beat (Cooper *et al* 1999, James 1991). Normal blood pressure is measured as less than 120 mmHg systolic and less than 80 mmHg diastolic; hypertension is measured as having 140-180 mmHg systolic or 80-110 mm Hg diastolic (Understanding Blood Pressure Readings 2017). Accurate blood pressure measurements are crucial for hypertension management (Ruhman 2012). Blood pressure, in population studies, is generally measured in the brachial artery (James 1991). The problem with blood pressure arises when there is too much pressure on the heart to pump blood throughout the body. This pressure causes the heart's walls to weaken and often causes side effects such as lack of oxygen to different parts of the body, the brain in particular; this is why untreated high blood pressure often results in stroke (James and Brown 1997). Hypertension is a major risk factor for other diseases as well. Some of these diseases include congestive heart failure, kidney failure, renal insufficiency,

coronary heart disease, and other diseases that affect the heart and blood vessels (Young *et al* 2005, Weinberger 1996, Burt *et al* 1995, Saunders 1991). Blood pressure is thought to be influenced by factors such as race, age, sex, environment, social situations, and genetics (James and Brown 1997, James and Pecker 1994).

Several studies have shown that African Americans are more likely to suffer from high blood pressure than white Americans (O'shaughnessy 2006, Wolf-Maier 2003, Cooper *et al* 1999, Cooper *et al* 1997, Weinberger 1996, Saunders 1991). While the percentage of white men and women who suffer from hypertension range around 31% and 28%, African American men and women suffer from substantially higher rates at 42% and 44% (Rahman 2012). As well as being highly prevalent in the African American community, hypertension is also a major cause of death, killing twice as many blacks as whites (Cooper *et al* 1999). There is a variety of speculation about the cause of this phenomenon. Some of the more popular speculations include increased salt intake and environmental stress (Rahman 2012, Jackson 2006, Young *et al* 2005, Cooper *et al* 1997, Weinberger 1996, Jackson 1991, Wilson and Grim 1991). However, it is important to note that data that support the narrative that blacks suffer from higher than normal rates of hypertension come from studies that focused on black and white comparisons of hypertension prevalence (Cooper *et al* 1997). One comparative hypertension study looked at African Americans, Africans, and Caribbeans risk susceptibility (Cooper *et al* 1997). Cooper *et al* conducted door to door surveys in each respective community while also taking blood pressures and collecting anthropometric measurements. They found that sodium excretion was lowest in Africa, highest in the United States, and the Caribbean fell somewhere in-between (Cooper *et al* 1997). They also found that Nigerians had the

lowest percentage of hypertensive individuals compared to the other communities studied; in contrast, African Americans in the United States had the highest percentage (Cooper *et al* 1997). Their results allowed them to make linkages between social and cultural factors and hypertension among those of African descent, further strengthening the idea of environmental factors as primary determinants in hypertension susceptibility. However, a recent report in 2013 states that Africans now have higher average blood pressure than their counterparts in Europe and the United States, which is cause for further investigation (De *et al* 2013).

Causation and Genetic factors:

Cooper *et al.* suggest forgoing the idea that race is the sole factor of high hypertension rates in African Americans (1999). In fact, they suggest holding constant genetic backgrounds of people in distinct environments in order to focus on environmental and behavioral variation (Cooper *et al* 1999). This is the best solution to them because they, as well as other studies, acknowledge the fact that hypertension arises through many factors and complex interactions rather than being the result of one thing (Cooper *et al* 1999, Weinberger 1996). Blood pressure is thought to be influenced by factors such as race, age, sex, environment, social situations, and genetics as well (James and Brown 1997, Weinberger 1996, James and Pecker 1994, Jackson 1991). Stress and lived experiences, such as discrimination are also thought to influence high blood pressure (Hicken *et al* 2014, Spruill 2010, Esler *et al* 2008,Willams *et al* 2003). Hypertension displays a wide variation in severity genotypically and phenotypically (Jackson 1991). In fact, Jackson names genetic recombination, genetic drift, gene flow,

and natural selection as traditional factors that could influence this pattern (1991). Other biologists and geneticists also recognize how these factors can play a role in adaptation (Tishkoff 2003, Campbell and Tishkoff 2008). Climate adaptation has been a prominent factor in discussing Africans susceptibility to hypertension (Jackson 2006, Young *et al* 2005, Jackson 1991, Wilson and Grim 1991).

Salt sensitivity has been used as a primary method for showing hypertension as being more prevalent in African Americans than other races (Weinberger 1996). Salt sensitivity is how one reacts to salt intake. Although the criteria, definitions, and overall methodology of salt intake measured in humans has been vast, there has been a general consensus in regard to observations (Weinberger 1996, Sullivan 1991). Overlack and colleagues (1993) defined salt sensitivity as a significant rise in blood pressure when individuals switch from a low to a high sodium chloride intake while Sullivan (1991) defined salt sensitivity as an increase in blood pressure that results in a relatively high sodium intake. However, some studies choose not to define salt sensitivity at all. The relationship between salt intake and blood pressure has been recognized since the 1960s (O'shaughnessy *et al* 2006). By looking at hypertension prevalence in 5 five geographically separate areas, Lewis Dahl was first to propose that blood pressure rises in direct correlation with salt consumption (Dahl 1960). According to Elliot and colleagues (1996) while African Americans have a higher than average risk of suffering from hypertension, larger studies have confirmed that overall, blood pressure among populations parallel that of increased salt consumption intake. There is still skepticism despite constant observations showing links between salt intake and blood pressure. Some of these skepticisms arise from the observation that not all individuals exhibit

changes in blood pressure after increased or decreased intake of salt (Weinberger 1996, Sullivan 1991).

Some researchers advocate that a family history of hypertension is linked with blood pressure salt sensitivity, insinuating that this phenomenon may be genetically inherited (Weinberger 1996). The salt hypothesis theory, by Wilson and Grim 1991, takes into account salt consumption as being a critical factor in determining hypertension but again states that African Americans are predisposed to have higher salt intake due to their transatlantic slave trade history, and therefore it is useful in evaluating effects of genetic and non-genetic factors related to causation in African Americans and other salt sensitive groups (Jackson 2006).

Biocultural approaches in providing alternative insights to interpretation:

There is plenty of anthropological research that examines environmental and cultural roles in human adaptation (Singer 1989, Livingstone 1958). Medical and biological anthropologists have formulated a biocultural model when looking at concepts of adaptability and human environmental interactions in order to better understand human genetic variation. Since Singer 1989, anthropologists have examined the role of culture and how it subsequently affects one's genes, coining the term "bio culture". Biocultural studies, according to McElroy are defined as:

> "research on questions of human biology and medical ecology that specifically includes social, cultural, or behavioral variables in the research design, offer valuable models for studying the interface between biological and cultural factors affecting human well-being (1990, abstract)."

Biocultural approaches, which focus on explaining variation as a response to larger components, have proven useful in generating holistic viewpoints on human biological variation (Dufour 2006). There are two biocultural approach models; one fuses biological, environmental, and cultural data while the more subdivided approach treats biological data as primary data and culture and environmental data as secondary (McElroy 1990). Research on human adaptability in various environments has been carried out (Baker 1969, Livingston 1958, Baker and Little 1976). An example of this comes from anthropologist Frank Livingstone. Livingstone linked population growth and subsistence strategy to sickle cell gene distribution. (Livingstone 1958).

While evidence of perpetual ethnic inequalities in health and healthcare in the United States is undeniable (Chapman 2005) approaches to combat this problem are still in the early stages. Anthropology has served as a tool for understanding ethnic health disparities by using ethnography to aid in new inequality knowledge for those who are affected as well as formulating alternative models, such as looking at pathogens bio-socially (Chapman 2005). Implementing anthropological methods, theory, and practice can supply different approaches and ways to interpret them to better understand social process and their role in ethnic health disparities (Chapman 2005).

Scholars have gone about studying hypertension in different ways. While some studies have used survey data to discuss hypertension's role in different populations (Wolf-Maier 2003, Cooper *et al* 1997), others have used strictly salt sensitivity analysis or gene analysis (Young *et al* 2005, Weinberger 1996, Elliot *et al* 1996, Overlack *et al* 1995, Overlack *et al* 1993, Sullivan 1991, Kawasaki *et al* 1978).

Chapter 3: Methods and Materials:

--Research Design

The study was conducted on the campus of Howard University April 11th 2017. The data used and analyzed was collected at a DNA screening workshop conducted at Howard University as well. The purpose of the workshop was to attract people from the African Diasporas to go through DNA screening and questionnaires to help build a robust database of African genomic variation.

The research design of the DNA collection consisted of two separate methodologies: genealogical survey data that emphasizes ancestry and health, and biological saliva samples. Both methodologies were needed in order to approach this database bioculturally. As previously stated, there are two separate ways to address biocultural approaches. One method fuses biological, environmental, and cultural data while the other more subdivided approach treats biological data as primary data and culture and environmental data as secondary (McElroy 1990).

--Population description and justification:

Africans and African Americans were the targeted population of this study. The purpose of this targeted sample is to draw on Africa's vast diversity and create a robust database that has the ability to address issues of contemporary health disparities while also taking into account African/African Americans' evolutionary history. Furthermore, focusing on individuals of African descent from the African diasporas allows researchers to address future research on what black-on-black comparisons reveal about environments and ancestry's role in causation in hypertension as well as to test the validity of salt sensitivity hypothesis.

-- Method development for making observations:

At the workshop, participants were first asked to complete a short three-part survey emphasizing ancestry and health. Part one of the survey asked ancestral questions concerning the origins of participants' maternal and paternal sides of the family. Part two dealt with family history of disease, while part three focused specifically on hypertension. After survey completion, participants were asked to contribute a sample of their saliva for the study. Saliva samples were collected by oral regurgitation of saliva into sterile test tubes; each test tube contained around 25 ml of saliva. The collected saliva samples will eventually be sequenced by the Genographic Project or Helix and samples that are not sequenced will be stored in a minus 80-degree C ultra-cold freezer and archived.

--coding of data

Coding of full survey:

Coding of the data took several steps. First, each section of the survey was coded separately on different excel files. Part 1 of the survey, the genealogical section was coded using alphanumeric coding.

Part 1 of the survey consisted of four questions that were asked of the participants. The four questions were: nationality, ethnicity, language, and place of birth. Participants then filled out their respected answers. Coding methods of all four questions were the same. For example, the first answer seen would equal 1, the next would equal 2 and so forth (00= no data, 01=American, 02=Caribbean). For both nationality and ethnicity numbers 00-31 represented general descriptions while 31 and continuing represented more specific answers for both paternal and maternal histories.

Part 2 of the survey which consisted of personal and family medical history was coded similarly. In this section participants checked whether or not they or anyone else in their family suffered from the given medical conditions: 1=yes and 2=no for this section.

Part 3 of the survey consisted of a hypertension questionnaire. The coding of this section was similar to the first two parts of the survey, but instead of the numbers corresponding with participants' answers they corresponded to pre-written responses (i.e. 1=low carbohydrate/sugar, 2=low cholesterol, 3=low salt, and 4=vegetarian).

Coding of barcodes for saliva samples (*Coding of first 36 bytes)*:

After the survey was coded barcodes were made for the test tubes. The barcodes consisted of 36 bytes and the universal barcode style of CODE 39 was used. The bytes were used as an additional code to know about the participant. The first 36 bytes of all three parts of the survey consisted of a code of easy identifiable factors for each participant. The order of the bytes was as follows: 1-2: database, 3-6:ID number, 7-8: gender, 9-18: latitude, 19-28: longitude, 29-30: current state, 31: current continent, 32-33: place of birth, 34: continent of birth, 35-36: database location.

Bytes 1 -2: Database

Sampling strategies used to obtain a representative subset were as follows: Participants were sorted into four different databases: CA (continental African), AD (African diaspora), RA (remote African), and AP (Arabian Peninsula). CA consisted of participants who had both parents and all 4 grandparents originated in Africa, while those sorted into AD consisted of participants who either had no parents originating in Africa or only one of the two. RA consisted of participants who had no African descent at all

(i.e. those from China, Germany, Pakistan etc.) Participants in AP consisted of people originating from countries in the Arabian Peninsula.

Bytes 3-6:ID number

Participants were given ID numbers at DNA collection workshop.

Bytes 7-8: Gender

Participants were identified by three different codes 01= male, 02=female, 00=no answer

Bytes 9-18: latitude and 19-28: longitude

Latitude and longitudes were calculated for the participants' place of birth. If participants did not disclose what city they were born in, the capital of the state/country they used was used to determine latitude and longitude. The purpose of getting exact locations of our participants is linked back to the lack of knowledge we have of African genetic variation in relation to geospatial ecology.

Bytes 29-30: current state and 31: current continent:

Participants' current state was determined by the address given in part three of the survey. Current state was the only section not coded alphanumerically, state codes were used.

Bytes 32-33: place of birth and 34: continent of birth

Participants' place of birth were determined by response in part one of the survey.

Bytes 35-36: database location

Participants' database location was determined by what database they were a part of (bytes 1-2). The purpose of this was to give more specific data on where participants came from. For example, CA database was split into more definite sections when coding for database location. WA (West Africa), EA (East Africa), SA (South Africa), CN (Central Africa), NA (North Africa).

--description of preliminary data analysis:

After the data were coded, data analysis of the database consisted of preliminary descriptive analysis to address these following questions:

1. How many individuals were in each database?
2. What were the gender distribution of the databases?
3. What were the nationalities and ethnicities of participants?
4. What are the age distributions of the databases?
5. What are the current residential status of participants? Which states are missing?
6. What are the place of birth of participants? What countries and geographical areas are missing?
7. What were the proportions of Howard University community members sampled?

Minitab and SPSS were used for primary data analysis. The coding of the survey data was Part one of the methodology. After saliva samples are sequenced they will be entered in the database allowing for further and complete analysis of a comprehensive biocultural database of Africans and African Americans.

Chapter 4: Selected Analysis Findings:

Gender, Age, and Numbers

Figure (1)

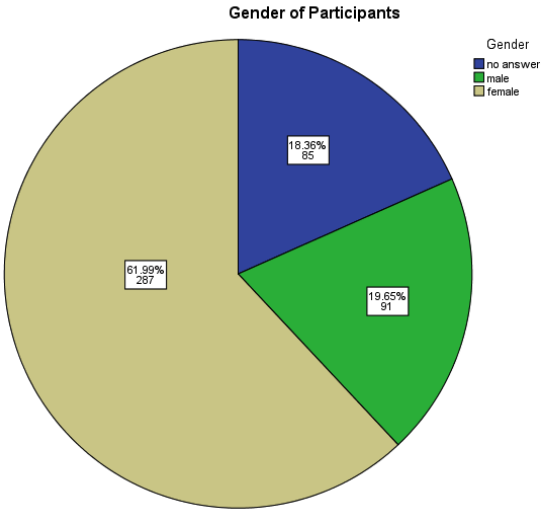

Figure (2)

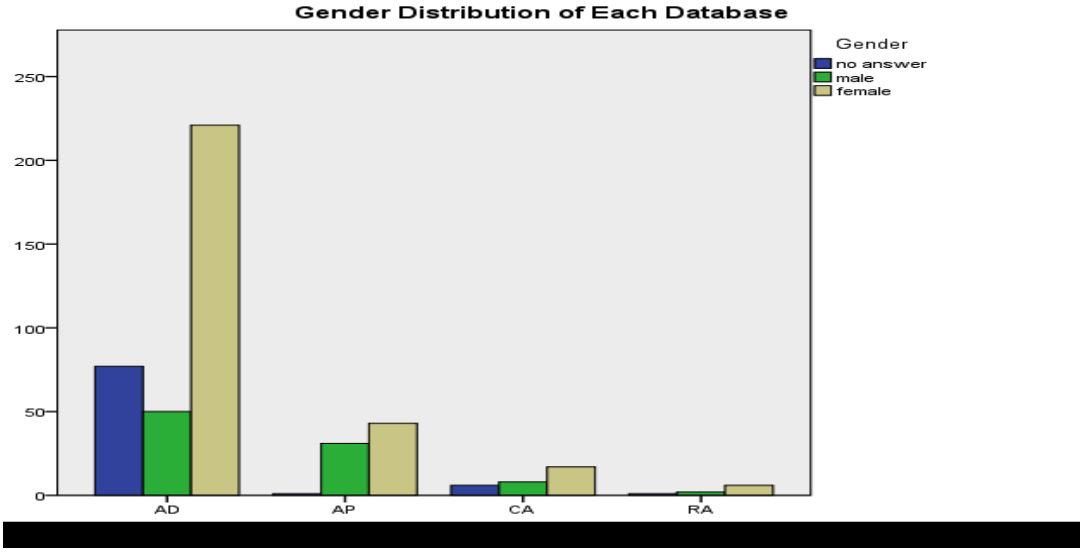

Figure (1) represents summary analysis of participants gender. In general, women participants accounted for 61.99% of all participants of the database, while men accounted for 19.65%. 18.36% of those surveyed reported no gender. More specifically, woman participation was higher than male in all four of the databases as seen in Figure (2).

Figure (3)

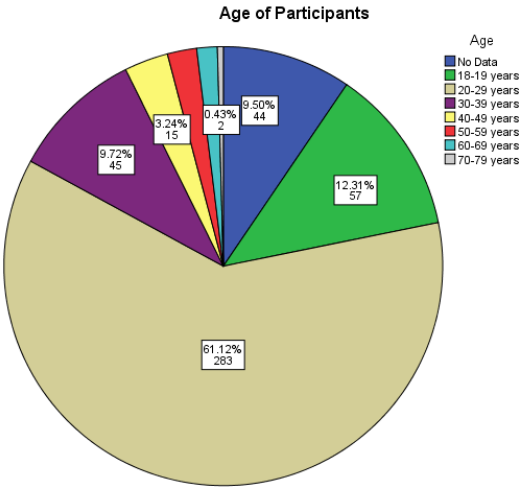

Figure (3.a)

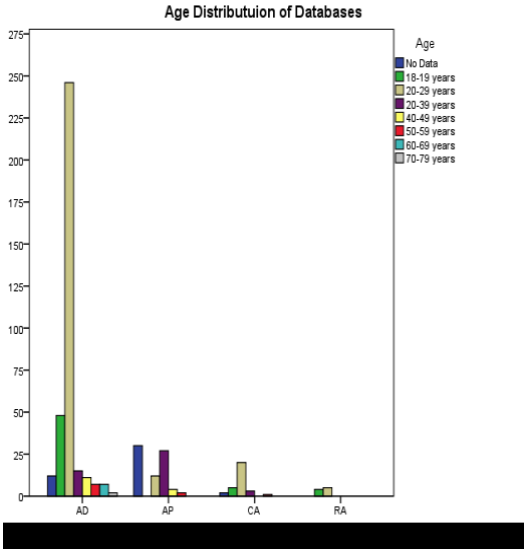

Figure (3) presents summary analysis of age distribution among participants. The majority of participants (61.12%) are between the ages of 20-29 years old. 9.72% of participants were between ages 20-39 while 3.24% ranged between ages 40-49. Participants 50 and over counted for less than 5% of all participants. 9.50% of participants did not specify their age. Figure (3.a) shows age distribution among each database. The AD database has the widest range of ages, while the RA database has a much narrower age range.

Figure (4)

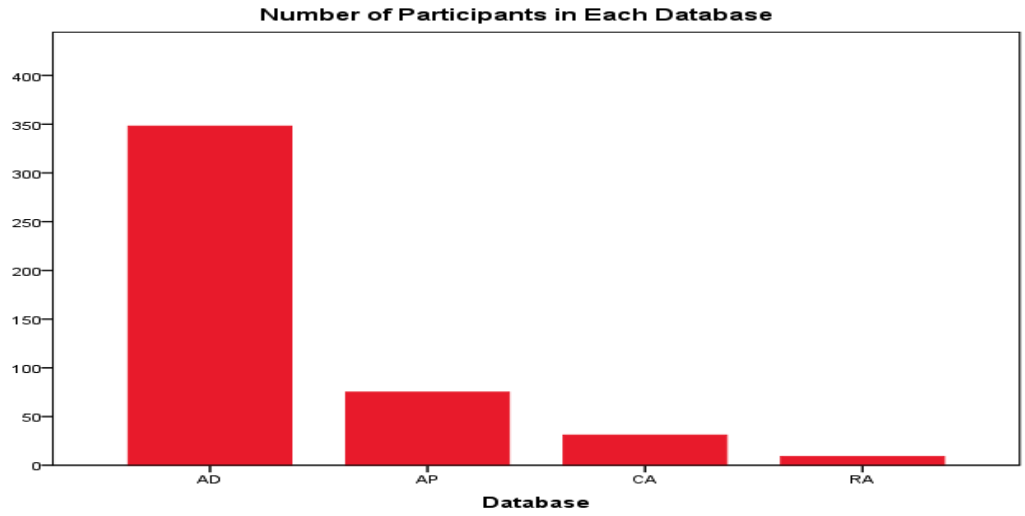

Figure (5)

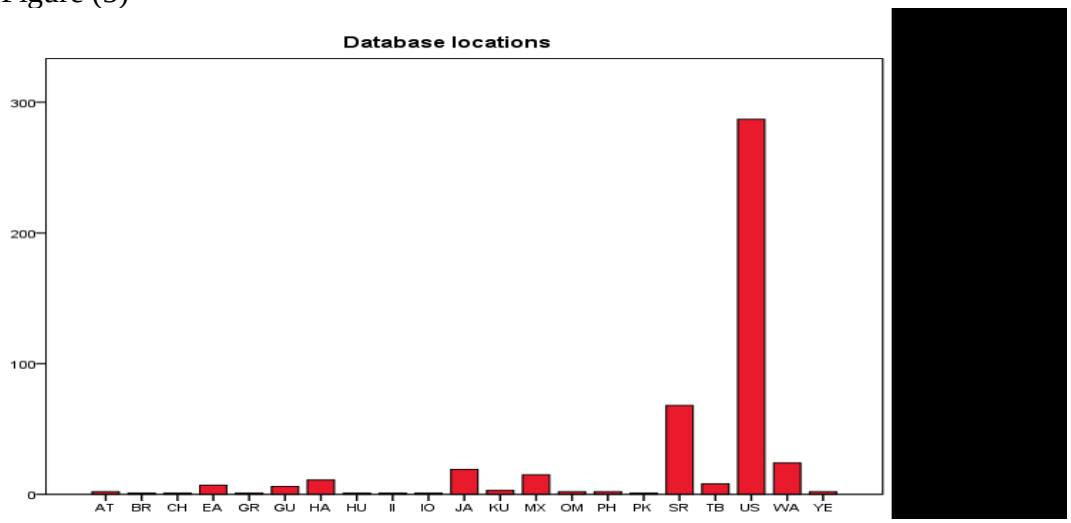

Figure (4) shows the number of participants in each database. In total, there were 463 participants. 348 (75.2%) in AD, 75 (16.2%) in AP, 31 (6.7%) in CA and 9 (2.0%) in RA. Based on figure (5) West Africa 77.4% (WA) and East African 22.6% (EA) make up all of database locations located in the CA database. North African 0% (NA) and South African 0% (SA) ethnic communities are absent.

Nationality, Ethnicity, and Geographical Distributions

Figure (6)

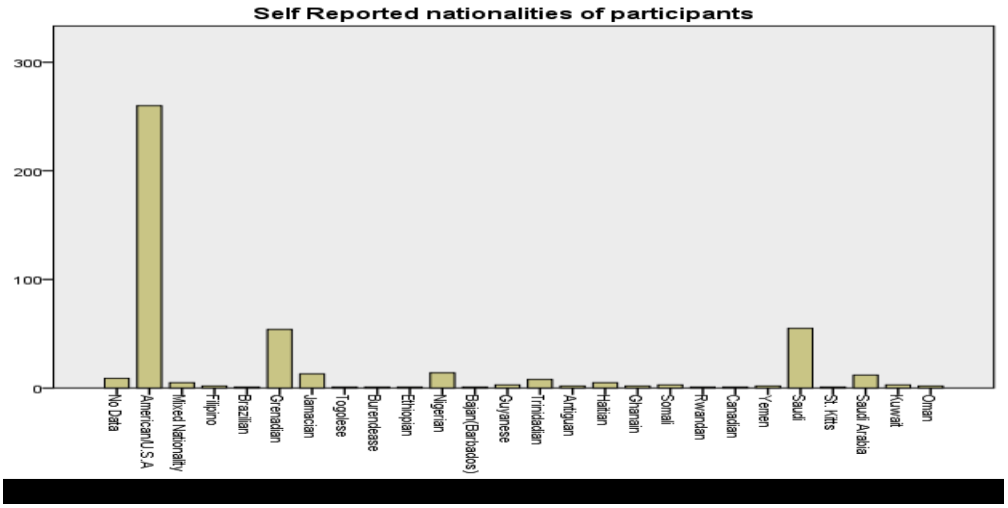

Figure (7)

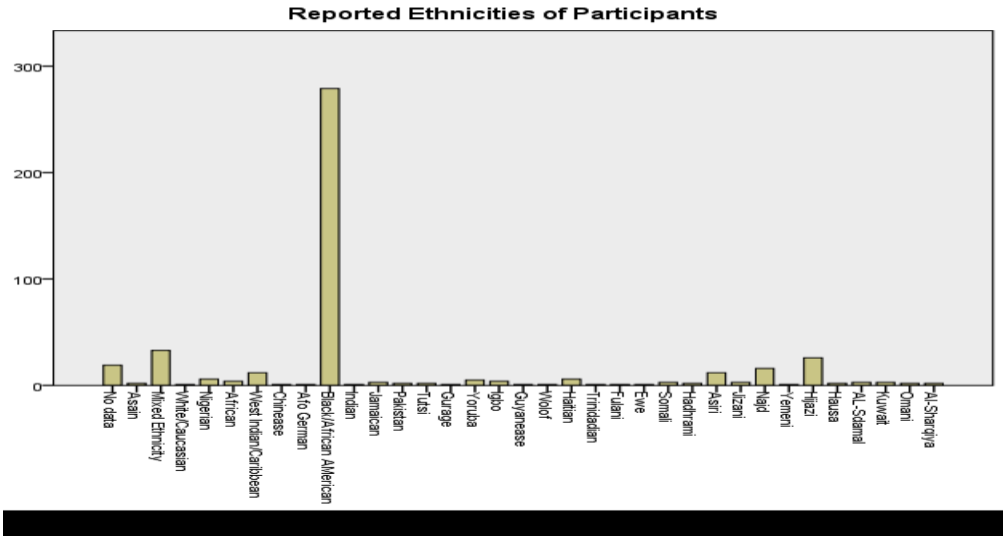

Figure (6) and Figure (7) show self-reported nationalities and ethnicities of participants of the database as a whole. There was a wide range in both nationalities (26) and ethnicities (35) reported. A majority of participants (56.2%) or about 260 participants self-reported their nationality as Black/African American. The percentage of self-reported Black/African American in terms of reported ethnicity increased to 60.3% or about 279 individuals. No participants self-reported themselves as "African" in terms of nationality. However, some participants self-identified themselves as Nigerian 14, Ethiopian 1, Togo 1, Burundi 2, Ghana 2, Somalia 3, Rwanda 1, making "Africans" account for 5.2% of self-reported nationalities. Self-reported "African" for ethnicity was much lower at .9%, after accounting for individuals who self-reported themselves as Nigerian 6, Igbo 4,

Yoruba 5, Somali 3, Tutsi 2, Gurage 1, Wolof 1, Fulani 1, and Ewe 1, “Africans” account for 6.0% of self-reported ethnicities.

Figure (8)

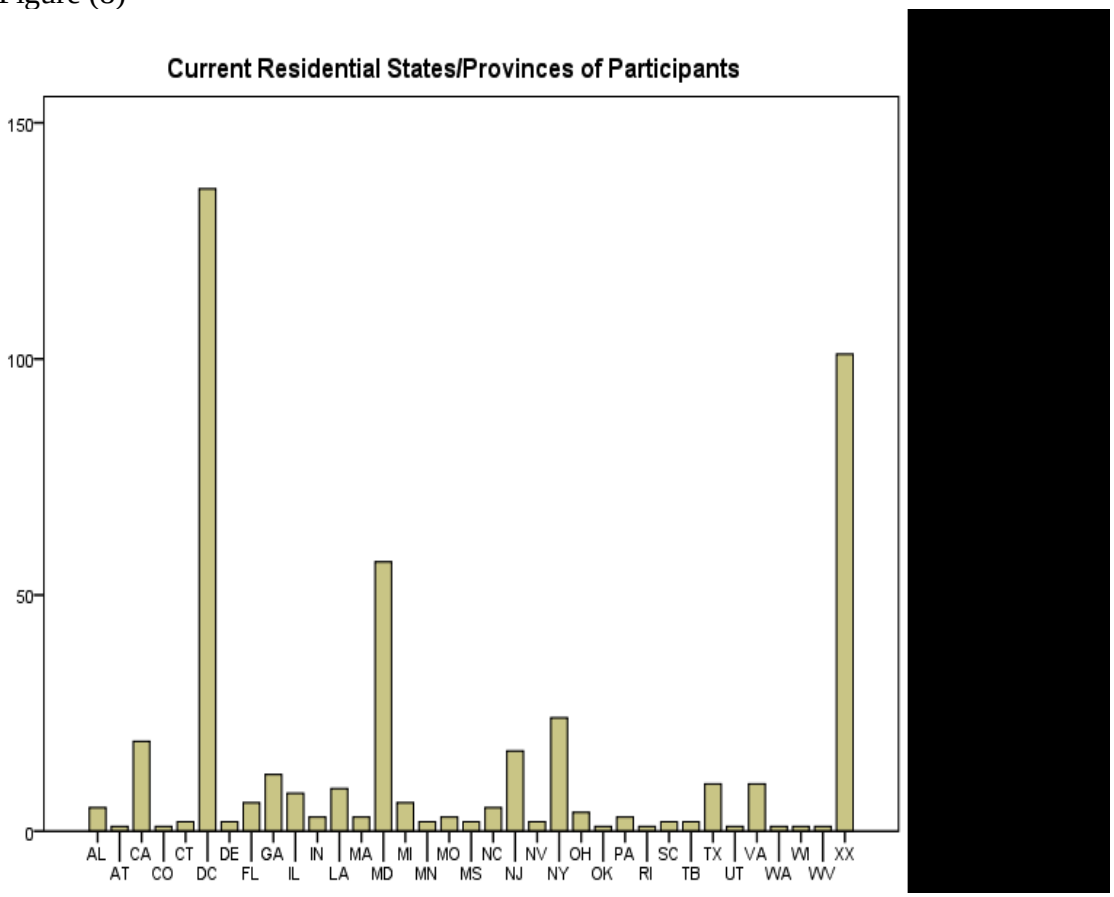

Figure (8) shows participants residential distribution. Most participants who reported where they were from are from D.C 29.4% and Maryland 12.3%. For a vast majority of participants current state/providence is unknown 21.8%.

Figure (9)

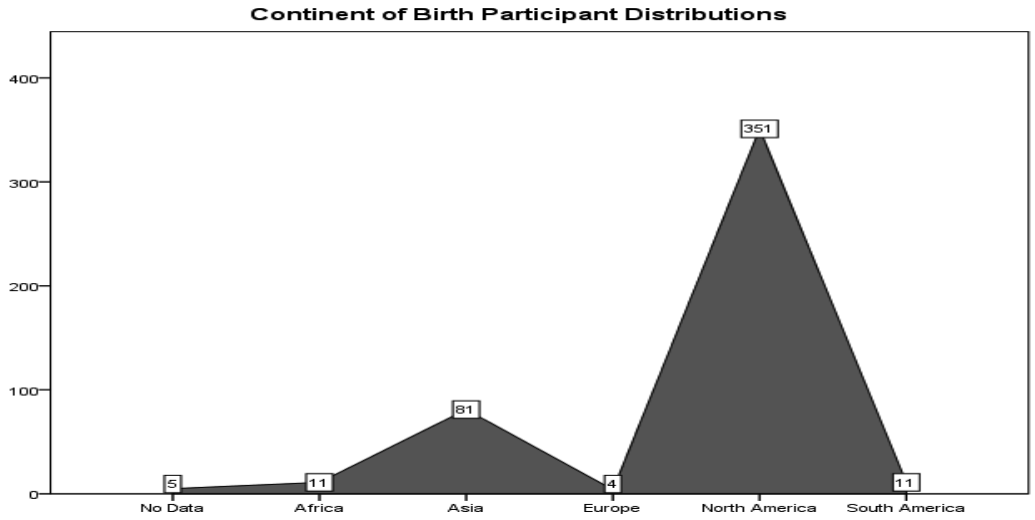

Figure (9.a)

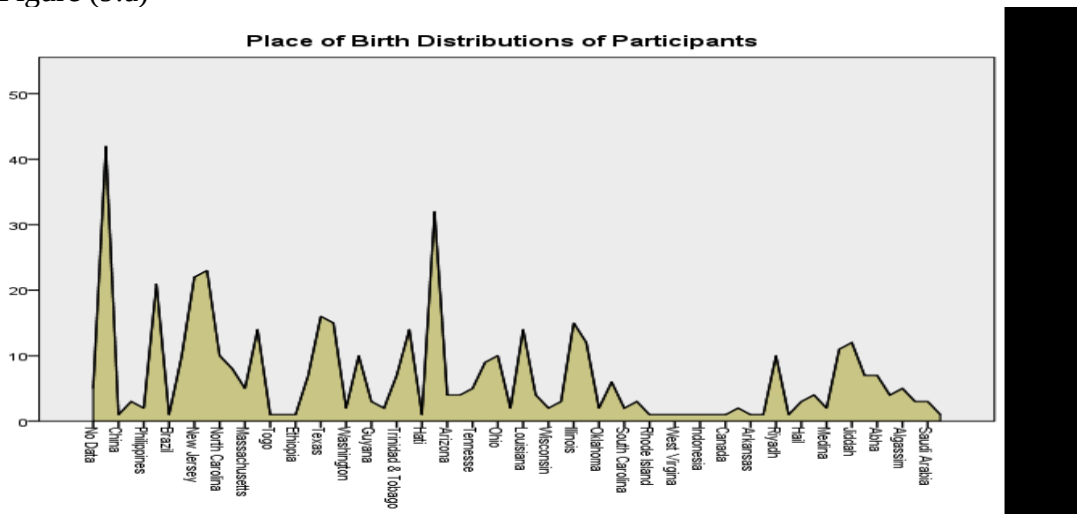

Figure (9.b)

Morocco	Somalia	Swaziland	Central African Republic
Algeria	Kenya	Lesotho	Chad
Tunisia (N=1)	Uganda	South Africa	Niger
Libya	Rwanda	Namibia	Mali
Egypt	Burundi	Botswana	Cameron
Sudan	Tanzania	Zambia	Benin
South Sudan	Malawi	Angola	Togo
Ethiopia (N=2)	Mozambique	Democratic Republic of Congo	
Eritrea	Madagascar	Congo	Cote d'ivoire
Djibouti	Zimbabwe	Gabon	Liberia
Sierra Leone	Guinea	Guinea-Bisseau	Gambia
Senegal	Cape Verde	Mauritania	Spanish Sahara

Figure (9) and figure (9.a) show continental distributions of participants. The majority 351 (75.8%) of participants were born in United States; while only 11 (2.4%) were born in Africa. Figure (9.a) shows a more in-depth figure of the country's participants were born in. Notable missing countries in Africa are listed below (figure 9.b).

Chapter 5: **Discussion:**

Bio-culturally informed African American centered databases are important for the advancement of human evolutionary history. Based on the preliminary data analysis, aspects that need to be addressed when constructing these databases are as follows:

Sex, Gender, and Numbers

Preliminary analysis of the database showed a heavy prevalence of women over men. Studies (Rahman 2012, high blood pressure 2016) have shown that, in general, men suffer from higher rates of higher blood pressure than women; however, looking strictly at African American communities, black women suffer disproportionately more than their black male counterparts (high blood pressure, 2016). The ratio of women to men in this database allows for potential comparative studies between African and African American women for further analysis relating to causation. However, obtaining more male participants is important in understanding what environmental, socioeconomic, and genetic factors play a role in women's hypertension rates.

According to the CDC around 11.1% of men and 6.8% of woman age ranged 20-34 suffer from high blood pressure (high blood pressure, 2016). The average age of database participants in our study ranged from 20-29 years old. While this is helpful and expected (because data was collected on a school campus), there needs to be a focus on collecting data from participants in the 35+ age range. In both men and women, there is more than a 50% chance of developing hypertension after the age of 55. Blacks develop hypertension more often and earlier than other ethnicities (high blood pressure, 2016).

This leads to database participation in general. Based on figures 4 and 5 participant involvement is heavily needed in the CA database, more specifically North, South, and East Africa. Without higher participation in this category comparative analysis will be difficult. In terms of diversity, getting a larger sample of Africans allows for comparative studies between African and African Americans subsets but also introspective research as well. Solutions to this include larger data expansion and outreach of African ethnicities. Addressing and understanding respective cultures plays a major role in gaining trust and participation from these ethnicities.

Nationality, Ethnicity, and Geographical Distributions

Figures 6 and 7 show the problems dealing with self-reported nationalities and ethnicities and the corrections that were made to address them. In the future, instead of allowing participants to write in their nationality or ethnicity, vast prerecorded categories should be available to allow for better analysis and easier coding for the database. Looking and understanding the role of self-identification is important because it plays a major role in precision medicine and treatment in general. There needs to be a methodology that allows for easier but mutually beneficial results for those giving data and those recording data.

Outreach is important in creating diverse collections of biological data. Diversity outreach is also important in reaching those not in the United States. Often, people associate African and African Americans as the same ethnic category, but with such

different environmental backgrounds understanding the difference between these different ethnic groups is important. Participant expansion is needed to diversify the database and allow for further comparative analysis. Getting more African participants has already been addressed, but getting more African participants directly from other parts of western and central Africa (not just Nigeria) will help give the database the versatility and range that a good comparative study would need for this particular study.

Bettering Survey

The ability to code a survey is essential to making data entry as easy as possible. A few ideas for making these steps easier is addressed throughout the next sections.

Conclusion

Stressing the importance and implementation of bio-culturally informed African American-centered databases is very important. I have addressed how these databases have the potential to reach the following four goals. One, highlight the diversity of African and African American communities through contextualized bio histories. Recognizing that African and African Americans are two separate populations is important in better understanding human evolution as a whole, especially in terms of human health and environmental correlations. Two, in recognizing these subset groups, databases have the ability to pave the way for precision medicine research and implementation. The future and advancement of precision medicine is pivotal in treatment and prevention in patients because it allows for personalized approaches to treatment based on environmental, lifestyle and gene variability; information easily attainable by databases. Three, building bio-culturally informed African-centered

databases will have the ability to eliminate health disparities seen in these populations, my paper focuses primarily on hypertension but there are a number of disparities that can be addressed such as breast cancer among African American women or diabetes. Using biocultural African centered-databases as direction, other minority groups can make similar databases to address health disparities among their communities. Fourth, I want to stress the idea of these databases being started or run by people who are in the given community. Being a part of the community gives the researcher the ability to not only know what questions to ask but to gain trust of those participants willing to help. Studies have proven that involvement of African American principle investigator's not only boosts participation of members but gives those involved less worry that their DNA/ results of the study will be used for motives that go against their benefit. The future and potential of bio-culturally informed African-centered databases is something too big to be ignored.

Bibliography:

1. American College of Medical Genetics, Storage of Genetic Materials Committee: ACMG Statement, Statement on storage and use of genetics materials. *American Journal Human Genetics. 1995*; 1499–1500.
2. Baharian, S., Barakatt, M., Gignoux, C. R., Shringarpure, S., Errington, J., Blot, W. J., Gravel, S. (2015). The Great Migration and African-American genomic diversity.
3. Baker, P. T. (1969). Human Adaptation to High Altitude. *Science,163*(3872), 1149-1156.
4. Baker T. P., Little A. P. (1976). Man in the Andes: multidisciplinary study of high-altitude Quechua. Stroudsburg, Pa.: [New York]: Dowden, Hutchinson & Ross; Exclusive distributor, Halsted Press.
5. Burchard EG. Missing patients. *Nature.* 2014 September; 513:301–302.
6. Burt, V. L., Cutler, J. A., Higgins, M., Horan, M. J., Labarthe, D., Whelton, P., Roccella, E. J. (1995). Trends in the Prevalence, Awareness, Treatment, and Control of Hypertension in the Adult US Population: Data from the Health Examination Surveys, 1960 to 1991. *Hypertension,26*(1), 60-69.
7. Bustamante CD, De La Vega FM, Burchard EG. Genomics for the world. *Nature.* 2011 July; 475:163–165.
8. Campbell, M. C., & Tishkoff, S. A. (2008). African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annual Review of Genomics and Human Genetics:* 9(1), 403-433.
9. Chapman, R. R. (2005). Radical contextualization: contributions to an anthropology of racial/ethnic health disparities. *Health:* 9(2), 145-167.
10. Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine.* 372(9), 793-795.
11. Corrigan, O., & Tutton, R. (2004). Genetic databases: Socio-ethical issues in the collection and use of DNA. Psychology Press.
12. Cooper, R. (1984). A note on the biologic concept of race and its application in epidemiologic research. *American Heart Journal,108*(3), 715-723.
13. Cooper, R., Rotimi, C., Ataman, S., Mcgee, D., Osotimehin, B., Kadiri, S., Wilks, R. (1997). The prevalence of hypertension in seven populations of west African origin. *American Journal of Public Health,87*(2), 160-168.
14. Cooper, R. S., Rotimi, C. N., & Ward, R. (1999). The Puzzle of Hypertension in African-Americans. *Scientific American,280*(2), 56-62.
15. Cozier, Y. (2004). Comparison of methods for collection of DNA samples by mail in the black womens health study. *Annals of Epidemiology:* 14(2), 117-122.
16. Crider, K. S. (2006). Racial and Ethnic Disparity in Participation in DNA Collection at the Atlanta Site of the National Birth Defects Prevention Study. *American Journal of Epidemiology*: 164(8), 805-812.
17. Crews DE and James GD (1991) Human evolution and the genetic epidemiology of chronic degenerative diseases. In GA Lasker and N Mascie-Taylor (eds.):

Application of Biological Anthropology to Human Affairs. London: Cambridge University Press, pp. 185-206.

18. Dahl, L. K. (1960). Possible role of salt intake in the development of essential hypertension. *Essential Hypertension*,53-65.
19. De, S. V., Akinyi, H., Oti, S., Olajide, A., Agyemang, C., Aboderin, I., & Kyobutungi, C. (2013). Status report on hypertension in Africa - Consultative review for the 6th Session of the African Union Conference of Ministers of Health on NCD'ss. *Pan African Medical Journal,16*.
20. Deschênes M, Cardunal G, Knoppers BM, Glass KC (2001): Human genetic research, DNA banking and consent: a question of "form"? *Clinical Genetics:* 59: 221–239.
21. Diaz, V. A., Mainous, A. G., McCall, A. A., & Geesey, M. E. (2008). Factors affecting research participation in African American college students. *Family Medicine-Kansas City*: 40(1), 46.
22. Dufour, D. L. (2005). Biocultural Approaches in Human Biology. *American Journal of Human Biology*: 18(1), 1-9.
23. Elliott P, Stamler S, Nichols R. (2006) for the Intersalt Cooperative Research Group. Intersalt revisited: further analyses of 24 hour sodium excretion and blood pressure within and across populations.
24. Esler M, Eikelis N, Schlaich M, Lambert G, Alvarenga M, Dawood T, Kaye D, Barton D, Pier C, Guo L, Brenchley C, Jennings G, Lambert E. (2008). Chronic mental stress is a cause of essential hypertension: presence of biological markers of stress. Clinical and Experimental Pharmacology and Physiology 35: 498 –502.
25. Godard, B., Schmidtke, J., Cassiman, J., & Aymé, S. (2003). Data storage and DNA banking for biomedical research: informed consent, confidentiality, quality issues, ownership, return of benefits. A professional perspective. *European Journal of Human Genetics*,11.
26. Hicken, M. T., Lee, H., Morenoff, J., House, J. S., & Williams, D. R. (2014). Racial/Ethnic Disparities in Hypertension Prevalence: Reconsidering the Role of Chronic Stress. *American journal of public health*, *104*(1), 117.
27. Jackson, F. L. (1991). An evolutionary perspective on salt, hypertension, and human genetic variability. *Hypertension,17*(1).
28. Jackson, F. L. (2006). Anthropological Science and the Salt-Hypertension Hypothesis. *Transforming Anthropology,14*(2), 173-175.
29. James, G. D. (1991). Blood pressure response to the daily stressors of urban environments: Methodology, basic concepts, and significance. *American Journal of Physical Anthropology*: 34(S13), 189-210.
30. James, G. D., Brown, D. E. (1997). The Biological Stress Response and Lifestyle: Catecholamines and Blood Pressure. *Annual Review of Anthropology,26*(1), 313-335.
31. James GD, Pecker MS. (1994). *Aging and blood pressure.* In *Biological Anthropology and Aging:Perspectives on Human Variation Over the Life Span*, ed. DE Crews, RM Garruto, pp. 321–38. Oxford: Oxford Univ. Press

32. Jameson, J. L., & Longo, D. L. (2015). Precision medicine—personalized, problematic, and promising. *Obstetrical and Gynecological Survey*: 70(10), 612-614.
33. Jenkins, M. M., Reed-Gross, E., Rasmussen, S. A., Barfield, W. D., Prue, C. E., Gallagher, M. L., & Honein, M. A. (2009). Maternal attitudes toward DNA collection for gene-environment studies: A qualitative research study. *American Journal of Medical Genetics Part A*: 149A(11), 2378-2386.
34. Kawasaki, T., Delea, C. S., Bartter, F. C., & Smith, H. (1978). The effect of high-sodium and low-sodium intakes on blood pressure and other related variables in human subjects with idiopathic hypertension. *The American Journal of Medicine,64*(2), 193-198.
35. Livingstone FB. 1958. Anthropological implications of sickle cell gene distribution in West Africa. *American Anthropology:* 60:533–562.
36. Martin P, Kaye J: The use of biological sample collections and personal medical information in human genetics research. London: The Welcome Trust; 1999.
37. McElroy, A. (1990). Biocultural models in studies of human health and adaptation. *Medical Anthropology Quarterly:* 4(3), 243-265.
38. McEwen JE, Reilly PR: Setting standards for DNA banks: toward a model code of conduct, Microb. Comp. Genomics 1996; 1: 165–177.
39. McQuillan G, Pan Q, Porter KS. Consent for genetic research in a general health population: An update on the National Health and Nutrition Examination Survey experience. Genetics in Medicine. 2006;8(6):354–360
40. Medical Research Council: Human Tissue and Biological Samples for Use in Research, Operational and Ethical Guidelines, 2001.
41. Moutel G, de Montgolfier S, Meningaud JP, Herve C: Bio-libraries and DNA storage: assessment of patient perception of information. Med Law (2001); 20: 193–204.
42. National Bioethical Advisory Commission (NBAC): Research involving human biological materials: ethical issues and policy guidance. Rockville, MD; August 1999
43. O'shaughnessy, K. M., & Karet, F. E. (2006). Salt Handling and Hypertension. *Annual Review of Nutrition: 26*(1), 343-365.
44. Overlack, A., Ruppert, M., Kolloch, R., Gobel, B., Kraft, K., Diehl, J., Stumpe, K. O. (1993). Divergent hemodynamic and hormonal responses to varying salt intake in normotensive subjects. *Hypertension: 22*(3), 331-338.
45. Overlack, A., Ruppert, M., Kolloch, R., Kraft, K., & Stumpe, K. O. (1995). Age is a major determinant of the divergent blood pressure responses to varying salt intake in essential hypertension. *American Journal of Hypertension,8*(8), 829-836.
46. Rahman, M. (2012). Hypertension in African Americans. *Diabetes and Hypertension,* 25-34.
47. Reed, F. A., & Tishkoff, S. A. (2006). African human diversity, origins and migrations. *Current Opinion in Genetics & Development:* 16(6), 597-605.
48. Reilly PR: Efforts to regulate the collection and use of genetic information. Arch Pathol Lab Med 1999; 123: 1066–1070.

49. Saunders, E. (1991). Hypertension in African-Americans. *Circulation,83*(4), 1465-1467.
50. Schlebusch, C. M., Skoglund, P., Sjodin, P., Gattepaille, L. M., Hernandez, D., Jay, F., Sen, Li., De Jongh. M., Singleton, A., Blum, G.B. M., Soodyall, H., Jakobsson, M. (2012). Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History. *Science,338*(6105), 374-379.
51. Schwartz, Marc D. "Consent to the use of stored DNA for genetics research: a survey of attitudes in the Jewish population." *American Journal of Medical Genetics Part A* 98.4 (2001): 336-342.
52. Shaibu, S. (2007). Ethical and cultural considerations in informed consent in Botswana. *Nursing ethics*, *14*(4), 503-509.
53. Singer, M. (1989). The limitations of medical ecology: The concept of adaptation in the context of social stratification and social transformation. *Medical Anthropology:* 10(4), 223-234.
54. Sirugo, G., Hennig, B. J., Adeyemo, A. A., Matimba, A., Newport, M. J., Ibrahim, M. E., Williams, S. M. (2008). Genetic studies of African populations: an overview on disease susceptibility and response to vaccines and therapeutics. *Human Genetics:* 123(6), 557-598.
55. Spruill, T. M. (2010). Chronic Psychosocial Stress and Hypertension. *Current hypertension reports*, *12*(1), 10-16.
56. Stoneking, M., Fontius, J. J., Clifford, S. L., Soodyall, H., Arcot, S. S., Saha, N., Jenkins, T., Tahir, A. M., Deininger, L. P., Batzer, M. A. (1997). AluInsertion Polymorphisms and Human Evolution: Evidence for a Larger Population Size in Africa. *Genome Research,7*(11), 1061-1071.
57. Sullivan, J. M. (1991). Salt sensitivity. Definition, conception, methodology, and long-term issues. *Hypertension,17*(1).
58. Tishkoff, S. A., & Verrelli, B. C. (2003). Patterns of Human Genetic Diversity: Implications for Human Evolutionary History and Disease. *Annual Review of Genomics and Human Genetics:* 4(1), 293-340.
59. Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., Williams, S. M. (2009). The Genetic Structure and History of Africans and African Americans. *Science:* 324(5930), 1035-1044.
60. Weinberger, M. H. (1996). Salt Sensitivity of Blood Pressure in Humans. *Hypertension,27*(3), 481-490.
61. White, M. (2016, July 9). How Slavery Changed the DNA Of African Americans. *Pacific Standard*.
62. Williams, D. R., Neighbors, H. W., & Jackson, J. S. (2003). Racial/ethnic discrimination and health: Findings from community studies. *American Journal of Public Health, 93*(2), 200-208.
63. Wilson, T. W., & Grim, C. E. (1991). Biohistory of slavery and blood pressure differences in blacks today. A hypothesis. *Hypertension,17*(1).
64. Wolf-Maier, K. (2003). Hypertension Prevalence and Blood Pressure Levels in 6 European Countries, Canada, and the United States. *Jama,289*(18), 2363.

65. Young, J. H., Chang, Y. C., Kim, J. D., Chretien, J., Klag, M. J., Levine, M. A., Chakravarti, A. (2005). Differential Susceptibility to Hypertension Is Due to Selection during the Out-of-Africa Expansion. *PLoS Genetics: 1*(6).
66. Zimmer, C. (2016, May 27). Tales of African-American History Found in DNA. *The New York Times.* Retrieved June 25, 2017.