12-17-2014

Evaluating Population Origins and Interpretations of Identity: a Case Study of the Lemba of South Africa

Jessica R. Engel
Georgia State University

EVALUATING POPULATION ORIGINS AND INTEPRETATIONS OF IDENTITY: A CASE STUDY OF THE LEMBA OF SOUTH AFRICA

by

JESSICA ENGEL

Under the Direction of Dr. Bethany Turner-Livermore

ABSTRACT

This study compares genetics and linguistics of the Lemba, a population living primarily in South Africa, as a means to identify any possible correlation between these two sources, to better understand how identity is impacted by ancestry testing, and to examine the Lemba's claim to Jewish ancestry with this evidence. The methods compare allele frequency data from several populations that were expected, based on Spurdle and Jenkins (1996), Casanova et al (1985), Ritte et al. (1993), Santachiara Benerecetti et al (1993), and Soodyall (2013), to be geographically proximate to and thereby more closely related the Lemban people. Results were clustered by language community to detect possible correlations. The different frequencies considered yielded dissimilar relationships between genetic and linguistic clusters, thus supporting the independence of mechanisms of linguistic and genetic change. These results

contribute to the discussion of how identity can be validated or undermined by demonstrating three sources, geographic, linguistic, and genetic, by which to derive an identity and how these can produce contradictory answers.

INDEX WORDS: Population genetics, Linguistics, Identity, Lemba, Ancestry, Genetic testing

EVALUATING POPULATION ORIGINS AND INTEPRETATIONS OF IDENTITY: A CASE STUDY OF THE LEMBA OF SOUTH AFRICA

by

JESSICA ENGEL

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Masters of Arts in Anthropology

in the College of Arts and Sciences

Georgia State University

2014

EVALUATING POPULATION ORIGINS AND INTEPRETATIONS OF IDENTITY: A CASE STUDY OF THE LEMBA OF SOUTH AFRICA

by

JESSICA ENGEL

Committee Chair: Dr. Bethany Turner-Livermore

Committee: Dr. Steven Black

Dr. Isabel Mendizabal

Dr. Soojin Yi

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2014

DEDICATION

For my grandparents James L. Horak Sr. and Frances J. Horak who have taught me that learning is a lifelong endeavor and who have inspired me to be a scientist

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my advisor and committee chair, Dr. Bethany Turner-Livermore, for her guidance through this M.A. program and especially during the thesis research and writing process. I would also like to thank my committee members, Dr. Steven Black, Dr. Isabel Mendizabal, and Dr. Soojin Yi, whose work inspired me to pursue the major topics of this thesis.

I thank my fiancée, Giovanni Vargas, for his patience, for all the feedback he gave on my ideas and writing, for his continuous support of my career aspirations, and for helping me keep life in perspective.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 Purpose of the Study

This study aims to investigate any possible connections between genetic, linguistic, and geographical patterns of populations and to evaluate whether either type of data can be utilized to support the other. It also seeks to evaluate how identity is impacted by different methods of determining ancestry. The study focuses on a specific population known as the Lemba from Sub-Saharan Africa, an indigenous population who also claim Jewish ancestry, and re-considers this emic affiliation by consulting previous studies and analyzing a combination of linguistic and genetic data. In examining how ancestry tracing has impacted the Lemban community and its individuals' sense of identity, this study aims to contribute to the discussion of the validity of these genetic and linguistic sources that contribute to concepts of identity. Moreover, this study aims to inform the larger dialogue on how policy should be adjusted to accommodate the growing popularity and use of genetic ancestry tracing, as well as the accessibility of such testing products.

1.2 Expected Results

In my project, I consider the following questions:

1. Can one use linguistic groups as evolutionary units and is there enough evidence from population studies in genetics, geography, and linguistics to suggest that any of the three could be utilized to support the results of new studies from any of the others' types of data?
2. Specifically regarding the Lemba, do linguistic and genetic evidence support their claims to Jewish (i.e., European and Southwest Asian) heritage? What are the

foreseen enhancements/consequences of this affirmation/contradiction to their self-prescribed identity?

3. How does modern policy need to be shaped to handle human rights in the human genome age? How can this case population be utilized to enhance ethics surrounding population ancestry tracing and identity?

While genetic and linguistic changes occur under different conditions and at differing rates, it is still possible that a parallel could exist between these two types of data in a population since they can both be impacted by some shared factors, such as contact periods between two distinct groups. However, since the mechanisms that govern both sources are dissimilar, it is not expected that one source could be a predictor for the other but instead could provide supporting evidence for a historical event that would require further investigation. A clear understanding of how genetic, linguistic, and geographic data for a population relate would be a vital part when looking at how a community's identity is formed.

1.3 Overview of Chapters

Chapter 2 focuses on how scientists have attempted to study, quantify and categorize people as races in the past and also how the concept of race is perceived and applied in the modern day. It also briefly discusses more useful, mechanism-based models of human variation as well as more modern thinkers' contributions to the discussion of race as a social category. Chapter 3 first considers how genetics, linguistics, and geography have been used in comparison in the past. It briefly covers conditions and limitations for each type of source to be used in studying populations. Secondly, this chapter illustrates case studies in which one or more of these types of sources have been implemented for the sake of better understanding population origins and relationships between populations. It also highlights genome wide studies of larger

regions or continents. The survey of these case studies uses the population terminology set forth by the researchers to discuss their results.

Chapter 4 focuses on the impacts of genetic ancestry tracing on identity. A major part of this discussion is how different types of information about a person can shape an identity and how society perceives an individual or group's identity. It also offers recommendations for dealing with and preventing situations of identity conflict. Lastly, it provides two case examples of issues erupting in concepts of identity from genetic testing.

Chapter 5 describes the sample population, the Lemba, and the significance they hold to the conversation on genetic ancestry tracing and identity. Additionally, it briefly surveys Jewish history and explains where Lemba identity fits in this timeline. It also highlights the Lemba's neighboring groups of Bantu and Khoisan populations. This chapter also lays out the methodology utilized and describes the various genetic sources in the dataset. Chapter 6 examines the results produced from the methodology described in Chapter 5. It also provides figures displaying the allele frequencies for comparison between populations. The populations are grouped by language affiliation to allow for comparison of allele variant frequencies between language communities. The population labels utilized are those set forth by the original data source; these labels are often problematic, as will be discussed further.

Chapter 7 provides a discussion and interpretation of the results, discussing the patterns and exceptions to these patterns visible from the results. Furthermore, this chapter covers how the results compare to the hypothesis proposed and if the research questions were fully answered. This chapter also provides recommendations and directions for future studies and the broader significance of the results.

2 SCIENTIFIC ATTEMPTS TOWARDS HUMAN CATEGORIZATION

2.1 Historical Approaches to Classification

Classification has always been a foundational method employed by scientists as a means to simplify and better understand the complex nature of the world. Scientific curiosity over human biological variation gained momentum in the Age of Exploration while many scholars were still battling to sort out factual and fictional beings (Mielke et al. 2011). During this time, Europeans traveled by land and sea to areas of the world that they previously held no contact with and encounter peoples that they perceived to be drastically biologically and culturally dissimilar from themselves. Naturalists of this period began to attempt to classify and categorize other humans into new species, varieties, and types based on observational descriptions. The description by Francois Bernier in 1684 is believed to have been the first Eurocentric racial classification, separating humans into species of Europeans, Africans, Asians, and Lapps. The naturalists of the eighteenth century grappled with the idea that humans may be more similar to other animals than previously thought, despite this obvious conflict with religious traditional teachings. Many were conflicted with whether human diversity could be explained as separate species or as a spectrum of variety within a single species. Some scholars saw human diversity as the result of their surrounding environment while others proposed a trajectory of progress in which populations initially were savage, then barbaric, and ultimately civilized (Mielke et al. 2011).

The father of taxonomic nomenclature Carolus Linnaeus classified humans with primates but maintained their position at the top of the Great Chain of Being. He recognized a species as a unit that is immutable and a variety or subspecies as a unit that can exhibit unique characteristics, and grouped humans into subspecies of American, European, Asian, and African, with each group demonstrating specific traits. The American variety was “red, choleric, and upright” and

"ruled by customs", the European variety was "white, sanguine, and muscular" and "ruled by laws", the Asian variety was "pale yellow, melancholy, and stiff" and "ruled by opinions", and the African variety was "black, phlegmatic, and relaxed" and "ruled by caprice" (Mielke et al. 2011: 5-6). While these classifications did not specify a hierarchy it was relatively implied and provided a fair reflection of the Eurocentric perspective of superiority (Mielke et al. 2011).

The concept of race came into being with Johann Frederich Blumenbach, also known as the father of physical anthropology (Mielke et al. 2011). In his later works, he would coin the term "Caucasoid", based on the fair-skinned people living in the region near the Caucasus Mountains, and divide humans into five races: Caucasian, Mongolian, Ethiopian, American, and Malay. Furthermore, he established a model of what he deemed to be degeneration with Caucasians as the pinnacle from which the other groups deviated. He also studied skulls from his different groups extensively and proposed that environmental factors, such as lifestyle and customs, could impact morphology (Mielke et al. 2011).

Blumenbach would not be the last to develop his own racial grouping of humans. Much of these perspectives were highly ethnocentric and based on scholars' subjective observations. The hierarchy formulated from racial classifications stemmed issues of inequality and injustice, particularly as perceived inferior traits became justified as God's will. Still, some scholars argued for a single unified human species, a common origin, a recognition of the spectrum of diversity and the inability to draw boundaries between these prescribed human subspecies (Mielke et al. 2011).

The nineteenth century added mental and moral abilities to the discussion of race theory. The race concept fueled the debate between monogenesists, those supporting a single common origin of humans, and polygenesists, those supporting multiple Adams and origins for different

human groups, on human origins. Samuel Morton was a supporter of polygenesis, stemming from his research on mainly the size of human skulls. A trend of attempting to quantify and scientifically prove racial categories was primarily focused on the skull. From this endeavor, concepts of ideal humans or type specimens emerge and racism materializes and begins to take hold as a public mindset. Anthropometrics and anatomy were the tools scholars of this period implemented to establish specific requirements for each classification; each race was perceived to be limited to discrete traits rather than continuous variation (Mielke et al. 2011).

The twentieth century brought about the Modern Synthesis in evolutionary biology and provided new biochemical methods through which researchers continued their efforts toward a valid racial classification scheme, still isolating Europeans as distinctive from the rest of the world's populations. In the early twentieth century, the ABO blood antigen system was discovered as well as the findings of Gregor Mendel's early genetic research. This time period also initiated the scholarly conversation on the value of other factors playing a role in what makes people unique, such as social institutions, language, and religion. Race was deemed by some to be useless for describing or studying human variation; moreover, researchers such as Ashley Montagu recognized the important role that culture played in shaping what many scientists had attributed to biological variation between different groups. Consequently, the term ethnic group replaced the term race because it was more comprehensive in describing both biological and cultural aspects of group identity. Those scientists that supported this perspective did not disregard the clear visible differences between populations but believed that variation was continuous rather than discrete and pushed for a reconsideration of natural selection as a mechanism for impacting human diversity. Many new topics entered the realm of anthropological work from biology, including mechanisms influencing population changes

(Mielke et al. 2011: 13). The abundance of genetic material caused many scientists to reevaluate racial categories; while some furthered the idea that any differences would be minimal between racially defined groups, others saw genetics as a more objective way to determine races (Mielke et al. 2011).

Kroeber (1923) discussed race as strictly a biological concept rather than having a social application. If differences among people were based wholly on a single trait, it would be relatively simple to cluster individuals into groups. However, most classifications do not acknowledge the discrete or multi-faceted aspects of human biological traits, which make such groupings foundationally unfeasible. Plasticity, or the ability for the body to adapt and change according to its environment, is often disregarded, despite its seemingly rapid occurrence, as Franz Boas described in his seminal study of immigrants from Europe to the United States (Kroeber 1923, Boas 1912).

The 1950s brought an increased emphasis on the scientific method and studying dynamic processes to physical anthropology, compared to the more speculative and descriptive procedures that characterized earlier research into human variation and biological identities. However, racialist science still persisted; Garn and Coon, for example, proposed three tiers of races: geographical, local, and micro races (Mielke et al. 2011). The geographic races were based on the major continents, the local races were regional divisions of each continent, and the micro races were populations at the level at which breeding occur. The comparison of the five living races of Caucasoid, Mongoloid, Australoid, Congoid, and Capoid, as published by Carleton Coon in 1962, to the fossil record attracted a great deal of criticism over methodology and seemingly a return to the old ways of physical anthropology (Mielke et al. 2011). A rebuttal from Frank Livingstone came in 1962, where he argued compellingly that race had no place in natural

selection. He acknowledged human variation but recommended that scholars consider what he called clinal variation, that of a single trait across a geographic area. Many anthropologists began to speak out on the ambiguity and loaded nature of the term race (Mielke et al. 2011). One of the major figures in this discussion was Ashley Montagu, a pupil of Bronsilaw Malinowski, Franz Boas, and Margaret Mead, who was quoted on the propensity for anthropologists and biologists to try to categorize races: "In our own time valiant attempts have been made to pour new wine into the old bottles. The shape of the bottle, however, remains the same" (Montagu 1962:920). The succeeding decades of the 1960s and 1970s focused more on clines than races for studies of human diversity (Mielke et al. 2011).

Many disagree on whether racial classifications are fully removed from scientific inquiry today and whether these categories hold any value to research in human variation. A survey of published articles in the *American Journal of Physical Anthropology* found no real decline in the use of race from 1965 to 1996 (Mielke et al. 2011), but a follow up study (Mielke et al. 2011) did find this decline by 2000. Despite the race concept disappearing from scientific literature, it remains a component of the ideology held by many scientists (Mielke et al. 2011). Subjectivity is still evident in research studies, and biological traits are often misunderstood to relate to culture. It is therefore evident that attempts to use biological criteria, whether those criteria are phenotypic or genotypic in nature, are fraught with potential analytical and interpretive pitfalls (Molnar 1983).

.

3 PREVIOUS GENETIC AND LINGUISTIC STUDIES

3.1 Racialized traits and mechanisms of variation

Genes are not the sole or even optimal variables to use in defining a group of people, and many of the most commonly utilized traits used to distinguish between groups are malleable to environmental influence, of limited heritability, and complex in their inheritance. Skin color, for example, demonstrates some of the greatest variability in the human species and is polymorphic and environmentally plastic in its expression (Molnar 1983). This and many other polymorphic, non-concordant traits make humans an incredibly diverse species that cannot be usefully classified into biological races as scientists once hoped, and often still hope, to do (Molnar 1983). However, there are significant insights that can be gained in discussing population ancestry and its impacts on the construction of identity, using genetic data; this is particularly true when genetic data are not the sole variables used, but instead are analyzed with linguistic and historical data as well (Molnar 1983).

Each individual inherits genetic material from both his or her mother and father. Inheritance from either parent can be dominant or recessive, and consequently, there is a chance that such a trait may not appear or may appear in a form unlike is seen in the parents. A distribution and frequency of these traits across a population is where typological ideas of racial definitions arise. For this reason, neutral genetic markers are most frequently utilized for genetic

testing.

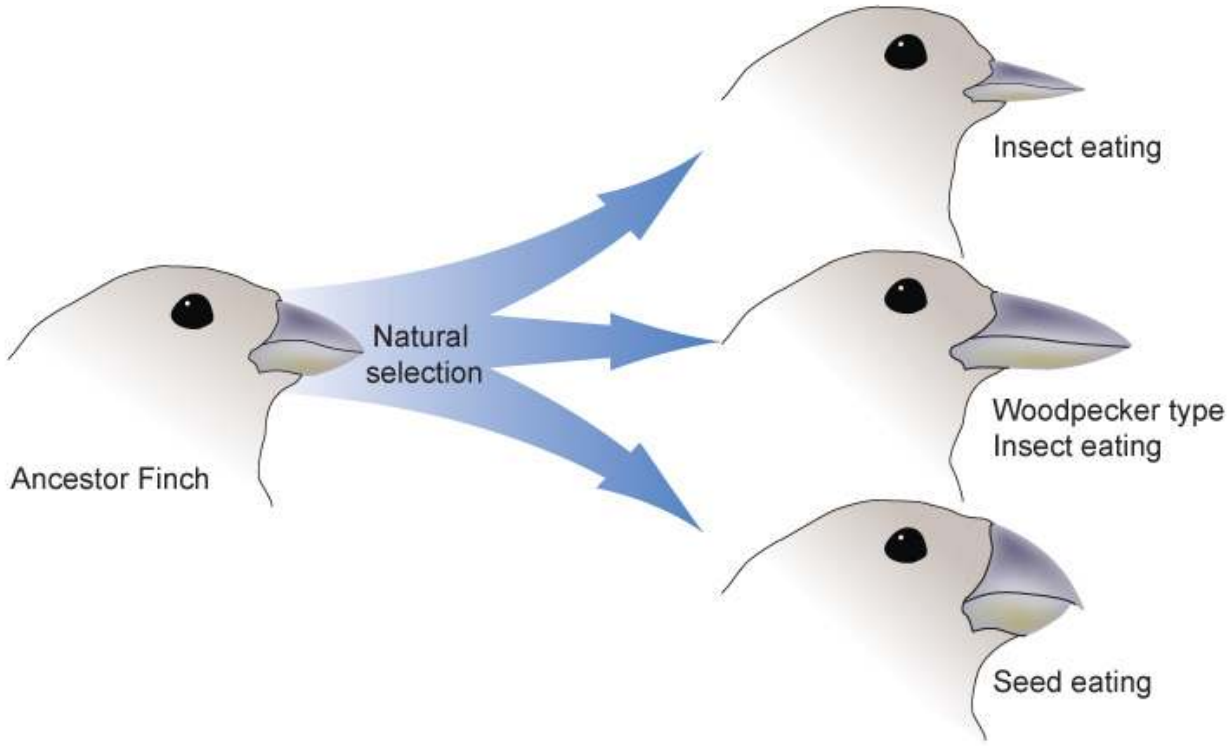

Figure 1, Natural Selection Example (Courtesy: National Human Genome Research Institute)

However, these frequencies can significantly change through four mechanisms or forces of evolution: mutation, natural selection, gene flow, and gene drift. Mutations can contribute completely new traits to a population; while the underlying cause of mutations can be environmental, the cause is not always known. Natural selection (see Figure 1) acts upon traits that are more or less advantageous, and accordingly, not at all possible genotypes are equally represented. Successful adaptations perpetuate the fitness of a species and are therefore more

likely to make it to the next generation.

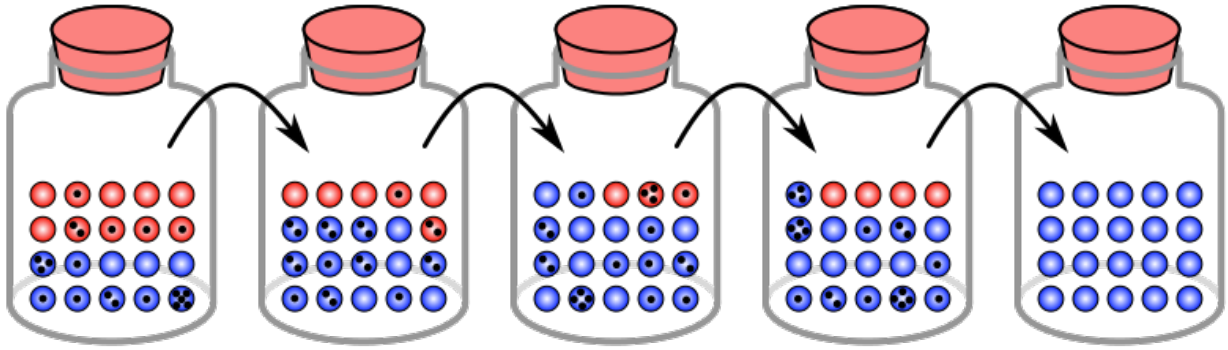

Figure 2, Genetic drift (Courtesy: Gringer, CC BY-SA 3.0)

Populations can also receive an influx of new genetic material from gene flow, or new genetic combinations that result from contact periods of migration, trade, or warfare. Genetic drift, or limitations of genetic variation being passed on due to population size (see Figure 2), can also eliminate potential genotypes from being passed to the next generation.

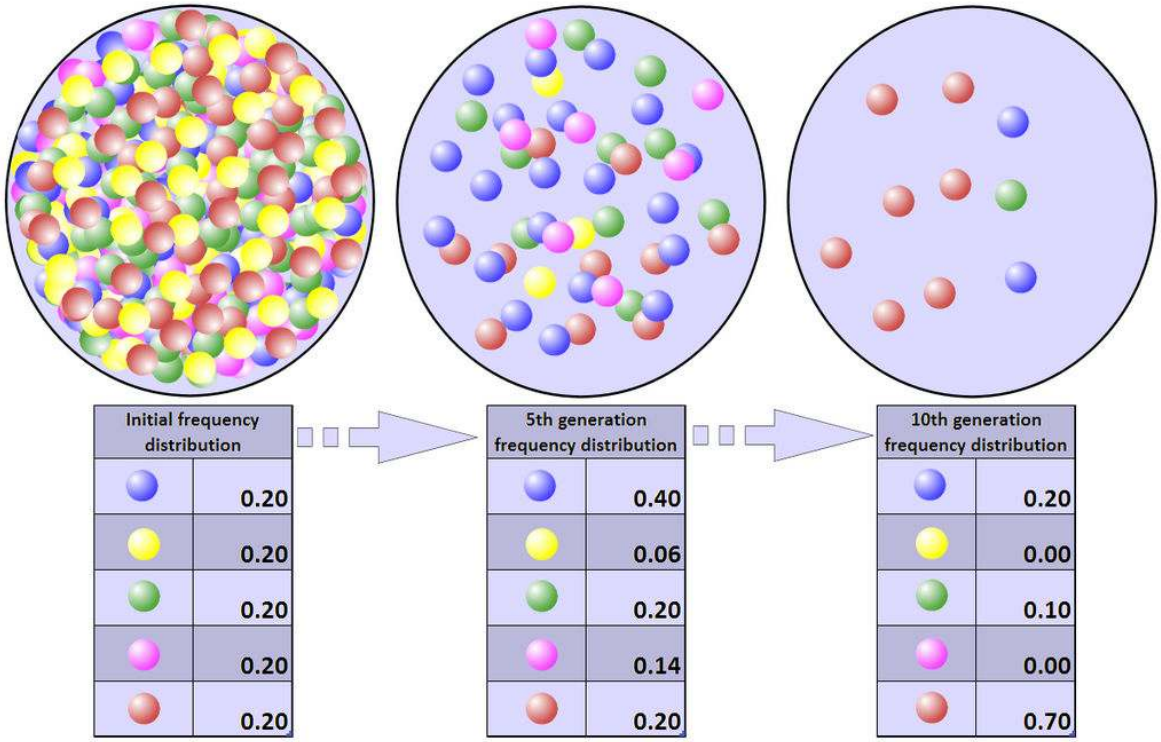

Figure 3, Bottleneck diagram (Courtesy: Professor Marginalia, CC BY-SA 3.0)

Bottlenecks and founder effects are examples of instances of genetic drift that result in reduced frequencies in particular alleles (see Figure 3 and 4) (Molnar 1983).

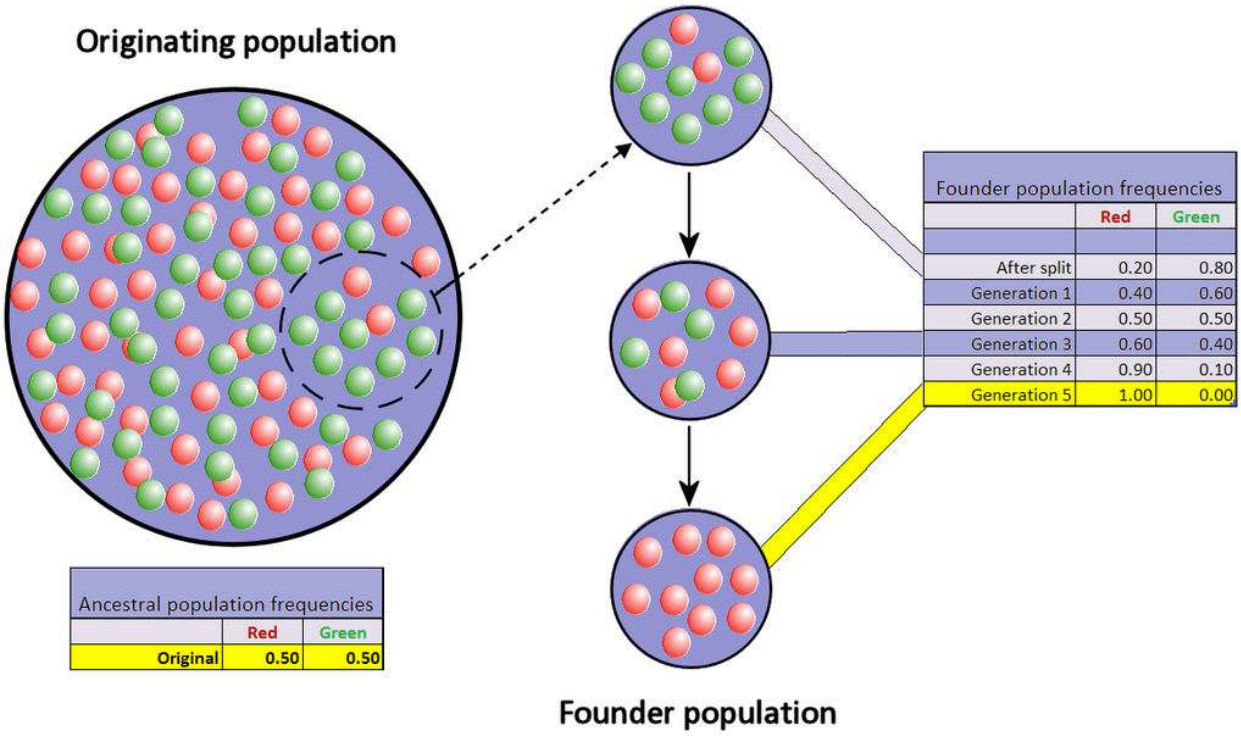

Figure 4, Founder effect diagram, (Courtesy: Professor Marginalia, CC BY-SA 3.0).

Population history is often studied using the following types of genetic data: mtDNA, the non-recombining portion of the Y chromosome (NRY), autosomal short tandem repeats (STRs), and single nucleotide polymorphism (SNP) microarrays. Mitochondrial DNA does not recombine and has a smaller effective population size, both making it easier to contrast between populations and reconstruct a phylogenetic tree. The NRY makes up roughly 95% of the Y chromosome. It also does not recombine and has a small effective population size. However, both are highly impacted by natural selection, have variable mutation rates, and reflect an exponentially less percentage the farther into evolutionary history that is considered. Short

tandem repeats are DNA sequences that have a variable number of short repeated segments, anywhere from 2-6 base pairs. Their high mutation rates make STRs advantageous for assessing more recent demographic events as well as examining closely related populations while making them restricted when studying older demographic events. SNPs can produce high-resolution results of population structure but are not as useful for establishing population divergences and shared origins from a parent population (Veeramah and Hammer 2014). Genetic diversity can be measured using different types of estimations, including the use of allele frequencies, the number of segregating sites, and the mean number of pairwise differences (Jobling 2004).

3.2 Processes of linguistic variation

In contrast to genetic change, language diversification does not occur at a constant rate; new dialects or daughter languages can manifest themselves somewhat unpredictably. Additionally, there are many possibilities for language extinction and shift, including periods of contact between two or more groups. Linguistic diversity does not correspond to any particular pattern but rather appears more randomly. Spread zones, an area where a small language group lives on a spatially wide area, are particularly interesting to study in linguistics with emphasis on dialects. We can expect that as groups moved from away from each other language components from either will become more distinct. Language spread occurs as a result of migration, expansion, or language shifts (Nichols 1997). Geographical barriers, such as mountains and coastlines, can also significantly impact language diversity. Geographic barriers restrict contact between groups on either side and therefore limit any flow of new linguistic features that could be shared. Additionally, this barrier stimulates an increase in diversification of language on either isolated side (Lee and Hasegawa 2014). The spread of political and economic entities also perpetuates the extinction of other languages (Nichols 1997).

Genetic and linguistic clusters must not be misunderstood to be synonymous but rather that language is a good indicator of an ethnic and genetic group. Therefore, it makes sense for genetically related groups to be similar linguistically. However, there are exceptions that are often attributed to more recent admixture between two populations and events of isolation. Both genes and language can be impacted by similar factors; change can initiate from an individual and spread throughout a population with either. Yet, such changes are far less frequent in genetics than in linguistic. Genetic changes are often in the form of mutations, which require direct transmission between related individuals and generational time. Linguistic changes, however, can occur much faster, especially because individuals do not have to be related and no time frame is required before passing it on. Humans are able to learn any language, but language can create barriers between groups that would otherwise exchange genetic material. As a result, it is most believable that language influences genes rather than the other way around (Cavalli-Sforza 2000).

3.3 Determining Population Ancestry

The study of population ancestry involves experts in many different disciplines, each of whom provide their own perspective and methodology to provide insights into the same questions. In recent years, anthropologists have illuminated notable correlations between genetic, linguistic, and geographic variables in populations in a number of regions worldwide. Much of the correlation present among genetic, linguistic, and geographic data can be attributed to changes in the population due to similar processes, such as geographic isolation or genetic exchange following a contact period, despite being different aspects of the population.

As a result of the many correlations identified between these factors, experts have attempted to infer a population's history, while also employing data from archaeology and

osteology (Barbujani 1997). A distinguished paper from Cavalli-Sforza et al. (1988) was one of the early attempts to synthesize the different data available to construct a more clear view of human evolution. In their phylogeny of the human species, the authors identified an initial split between Africans and non-Africans and a second split resulting in two major clusters of global populations. The first cluster defined by the authors included Caucasoids, East Asians, groups from Arctic regions, and Native Americans, while the second cluster included Southeast Asians, Pacific Islanders, New Guineans, and Australians. The authors found genetic distance to be in concordance with population divergence times as determined by the archaeological record and to intersect significantly with the linguistic families and superfamilies (Cavalli-Sforza et al. 1988).

Geographic distance and both physical and linguistic barriers limit the gene flow that will occur between multiple groups. Even still, linguistic and genetic diversity are variably impacted by isolation depending on the population's size. While a population's geographic history can be limited by the availability of written records or oral tradition, reconstructing both linguistic and genetic histories can help determine the origin of a population and shared origins with differing populations. In employing different methods for reconstructing origins, researchers can also attempt to identify points of splitting and divergence in language and genetic groups (Barbujani 1997).

The intertwined relationship between linguistics and genetics with regards to population history has encouraged some to speculate that either may be utilized to predict the other. Yet, the conclusion that a parallel evolution exists between linguistics and genetics still seems uncertain. Evolutionary processes can affect different types of genetic data uniquely and at dissimilar rates. When applying classic population genetics and nonparametric statistics to the relationship between linguistic and genetic populations, it is assumed that population separation occurred

instantly and without preceding intermixing. Additionally, differences in allele frequencies would have been the result of genetic drift. Under this assumption, these differences would appear more quickly in smaller groups than in larger groups (Barbujani 1997). With regards to language diversity, it can be assumed that linguistic diversification succeeds population separation or distinction because events of language replacement are notably rare, except in recent contexts resulting from colonial and post-colonial inequality (Ward et al. 1993). If operating from the assumption that a common language or related language points to a shared origin, one would expect to find parallels between markers of genetic and linguistic diversity. In scenarios where genetic differentiation from another population and departure from equilibrium occurs slowly, a separation or migration seems plausible as the geographical distance between the two groups would increase. In populations with a more abrupt change in allele frequencies, some form of isolation from the other population would be expected. Where both genetic and linguistic differentiation is remarkably stark simultaneously, barriers inhibiting free flowing reproduction between the two groups are most plausible, as they would drastically reduce the speed at which the population returns to genetic diversity equilibrium (Barbujani 1991).

Partnering linguistics and genetics in demographic studies does have its limitations. A major obstacle in these population ancestry studies is classification, whether genetic, cultural, or linguistic. Depending on the type of classification employed, different results could arise in the population's structure. As Nettle and Harriss (2003) point out, past studies of the relationship between population linguistics and genetics have often yielded results of strong correlation, making some believe that a parallelism between the two is a rule rather than an unusual feature. A linguistic tree implemented in some studies is derived from Ruhlen (1987); Nettle and Harriss (2003) have little confidence in the quality of this tree for scientific inquiry because while

roughly half of the tree is based on recognized language families the remaining half is classified based on plausible evidence but is ultimately speculative. A better understanding of how languages should be classified and how linguistic groups are related to each other will influence population ancestry studies. The same can be said for improving the construction of haplotype groups (Belle and Barbujani 2007). Another serious obstacle in this research is the privacy, proper use and consequences of genetic data for the participants. In projects where mass genetic surveying is being conducted, researchers often face accusations of being exploitative in either their interpretation of the data or what they do with the data after their study concludes (Bisol et al. 2008). This social aspect of genetic research will be discussed more thoroughly in the next section.

Additionally, it is difficult to be certain that the differences in language diversity are attributable to impacts of historical events or simply the population's geography (Barbujani 1991). While the influence of historical events and geography is sometimes considered together, too often geography alone is analyzed to be a correlate with genetic diversity (Bertorelle and Barbujani 1995). Shared components of language might suggest that two sedentary populations encountered and exchanged culture with each other, but it could also be indicative of populations with a language distinct from another moving into a new territory (Barbujani and Pilastro 1993). Furthermore, measuring language diversity can be variable depending on the type of linguistic data considered. Colonna et al. (2010) recommend utilizing syntactic data rather than lexicon data, which have been more commonly used in previous studies, because they are more easily comparable across different groups and are more stable and quantifiable. Just as challenging as it is to discern what causes these changes, researchers also disagree on what a correlation between linguistic and genetic factors truly means for our understanding of population histories. One

possibility is that their correlation is a reflection of the same events occurring. Another hypothesis is that they evolved independently but were impacted by many of the same influences (Nettle and Harriss 2003). While there is certainly justification for trying to identify a correlation between genetic and linguistic diversity, it cannot be disregarded that linguistic diversity does not necessarily exemplify an evolutionary unit as it undergoes processes that may be unique to it and dissimilar from processes affecting genetic diversity (Roychoudhury 2001). Other limitations to accuracy of results from studies of population structure include the quality of the database used, the number of genetic markers employed in the study, and the diversity present in the populations of the region (Shriver and Kittles 2004).

3.4 Case Studies

Recently, Cavalli-Sforza (2000) wrote a book tackling some of the major concerns in studies of population origins based on his own work. He discusses the comparison of genetic and linguistic trees and points out how groups can often be attributed the same label linguistically and genetically (Cavalli-Sforza 2000). An early study of population origins using genetics compares the categorized populations of Europeans, Africans, and Asians. Nei and Livshits found Europeans and Asians to be more closely related than Africans and deemed that their results supported the theory of an African origin for Homo sapiens. However, the authors used African Americans to represent African populations for roughly 17% of the loci. They offer that the European and African populations could be as similar as the European and Asian populations had they used individuals from Africa for all loci (Nei and Livshits 1989).

Rosenberg et al. (2002) studied 377 autosomal loci from 52 populations across the world to assess predefined population groups. They found that genotypes were more graduated than distinctive for each population and that clusters occurred frequently near geographical obstacles.

Additionally, they noticed that many populations that share a language also grouped into clusters based on genetics. Furthermore, populations that had been isolated for extended periods of history demonstrated relatively low heterozygosity, most likely due to genetic drift in smaller sample sizes (Rosenberg 2002).

In 1979, Carmelli and Cavalli-Sforza conducted a study of Jewish groups based on initially physical characteristics. They considered four blood markers for 12 Jewish groups and found when compared with other non-Jewish groups, their genes reflected their Middle Eastern origin and relatively minimal admixture with other groups. While similarities were observed with Near Eastern populations, the known gene frequencies for Asian populations was minimal at this time, and therefore, the authors did not draw conclusions from any correlations (Carmelli and Cavalli-Sforza 1979)

Genetic and linguistic diversity appears to follow similar patterns in European groups while a weaker relationship between the two is found within Native American, Asian, and North American populations (Monsalve et al. 1999). When comparing X chromosome diversity with language groups, a notable relationship was identified in only European and East and Central Asian populations (Belle and Barbujani 2007). In an analysis of worldwide samples from p49a,f/TaqI polymorphic marker of the Y chromosome, population structure seemed to parallel that of language families and was in agreement with archaeological evidence. Furthermore, geography correlated with this genetic marker as well. Poloni et al. found that this marker closely identified with the same patterns found in autosomal and mitochondrial DNA (Poloni et al. 1997).

A study by Nettle and Harris (2003) proposes that correlations between genetic and linguistic diversity only appear in particular circumstances (Nettle and Harriss 2003). Yet,

Cavalli-Sforza (1966) argues such interpretations are the result of the particular loci that have been selected for study and more research is needed to demonstrate true diversity through analysis of samples from numerous genome regions (Belle and Barbujani 2007). In a study by Belle and Barbujani, they respond to this need of further research by considering an extensive sample of polymorphic microsatellite loci that are well dispersed in the genome against language diversity. While they found there was a relationship between linguistic and genetic diversity, the authors propose that this could be explained by both being related to population geography. The study also investigates physical barriers and depicts the distinction of sharper contrasts in genetic data in populations that shared geographic borders but maintained languages from two distinct language phyla. This result expands upon the discussion of geographical barriers earlier in this chapter that physical barriers are not necessary to induce the same evolutionary effects of isolation; the same impacts can be derived from language barriers or distinct cultural, including religious, differences (Belle and Barbujani 2007).

Much of the genetic variation present in Europe is interpreted as a result of population expansion and structure during the Neolithic period and at the rise of agriculture (Barbujani and Pilastro 1993). As more food was able to be produced, Nostratic speaking and Near East dwelling populations were able to grow in mass waves and from these waves spouted new protolanguages, specifically Indo-European, Elamo-Dravidian, Afro-Asiatic, and Altaic; this concept is known as the Nostratic demic diffusion model, or the NDD model (Barbujani and Pilastro 1993). In Barbujani and Pilastro's reevaluation of the NDD model, they determined that the language clines present in Europe reflected the spread of agriculture and at least three of these protolanguages originating and spreading from the Near East into Europe (Barbujani and Pilastro 1993).

Correlations in European populations are of particular interest to those studying demography, specifically population structure and classification, as it is home to diverse languages and cultures. Yet, it is disputed whether language groups and genetic distances between European populations are really connected or if both are just a reflection of a shared geography. Even still, some groups that experience physical or cultural isolation, such as the Basque, Finnic, and Semitic language families, demonstrate particularly distant genetic relation to geographically nearby groups (Harding and Sokal 1988). These discrepancies in a possible relationship between the two could be attributed to the classification of language phylogenies in Europe (Harding and Sokal 1988). A subsequent study of European genetic and linguistic diversity found points of abrupt genetic change where physical and linguistic barriers were present. Barbujani and Sokal concluded that other factors that would isolate a population or cause them to migrate may have inhibited population admixture more than geographical distance (Barbujani and Sokal 1990).

A study by Ward et al. (1993) considered linguistic and genetic diversity against population geography in the Americas, where both have appeared as poor correlates in previous studies. Specifically, the authors studied three tribes, the Haida, the Nuu-Chah-Nulth, and the Bella Coola, in the Pacific Northwest, covering Amerind and Na-Dene, two language phyla, and Wakashan and Salishan, two Amerind language families (Ward et al. 1993). Their analysis of the groups' linguistics showed that the Haida, who spoke Na-Dene, diverged considerably more recent than the Nuu-Chah-Nulth and the Bella Coola. However, in studying their genetic differentiation, significantly less sequence divergence was present between the Haida and the two Amerind groups, than would be expected given the linguistic analysis. Ward et al. (1993) therefore proposed that linguistic and genetic diversity could occur at different rates; this is

particularly understandable since both will differentiate under unique processes. Furthermore, when language is understood as a part of culture, which would be influenced by social and historical events in a population, it is anticipated that language differentiation would not follow the static pace of molecular evolution but rather would be linked to instances of rapid change (Ward et al 1993), as is described earlier in this chapter.

A study by Gravel et al. (2013) sampled individuals from Colombia, Puerto Rico, and Mexico to study the diversity present in the Americas. The Americas represent an admixture occurring over time between African, European, and Pre-Colombian populations. They used these samples to investigate how people moved in the Americas from Eurasia through the Bering Strait and how they disperse after moving farther south. Their results reflect an initial bottleneck and then rapid divergence and migration after moving into the Americas (Gravel et al. 2013).

Africa poses an interesting area of study in population diversity; a third of the world's modern languages can be found spoken in its countries, the four language families being Niger-Kordofanian, Afroasiatic, Nilo-Saharan, and Khoisan. Additionally, continental Africa is believed to be the origins of the anatomically modern human population and the point from which all populations would diverge and migrate. For this reason, it is easy to understand how Africans demonstrate the highest genetic variation and the deepest lineages compared to non-Africans, when considering mitochondrial DNA, the non-recombinant portion of the Y chromosome, and autosomal DNA. In a study by Scheinfeldt et al. (2010), a correlation between linguistic and genetic distances for three of the four language families was reflected, with the Khoisan relationship lacking clarity (Scheinfeldt et al. 2010).

Another region noted for its extreme genetic diversity is the Caucasus area, between the Caspian and Black seas. This region seems to follow unique distribution patterns of linguistic

and genetic changes, which are distinct from those visible in other regions of Eurasia. While some have pointed to geographic subdivision as the reason for these distinct distributions, Barbujani et al. (1994) insist that the Caucasus area is too large for this to be the only factor. They propose several evolutionary scenarios that could explain the patterns that appear for the Caucasus area; however, the most probable scenario, considering the data sets is either a mass migration scenario or an elite dominance scenario. In a mass migration, members of an ancestral population would gradually separate, and this would impact the geography and genetic distance relationship (Barbujani 1994). In an elite dominance scenario, a new minority population would force their language upon an existing yet large population (Barbujani 1994). Overall, the Caucasus region does not seem to exhibit a correlation between language and genes to the extent of other regions. On the contrary, Barbujani et al. (1994) propose that the linguistic and genetic change may have occurred independently for most of the populations' history (Barbujani et al. 1994). A more recent study of the Caucasus region examined mitochondrial DNA and its relationship to linguistic families. The Caucasus region demonstrates less diversity than is found in the Near East but more diversity both within and between its populations than is found in Europe (Nasidze and Stoneking 2001). The authors' analysis of the mtDNA suggests that Caucasus populations may have been the result of admixture between European and Near Eastern groups or could have been ancestors to European populations (Nasidze and Stoneking 2001). Despite the presence of actual physical boundaries, such as the Caucasus Mountains, authors Nasidze and Stoneking found a stronger relationship between genetic diversity and geography than genetic diversity and linguistic diversity (Nasidze and Stoneking 2001).

An investigation of the population history of India has led to an examination of tribal groups where each speaks in a distinct language family, namely Austro-Asiatic, Dravidian, or

Tibeto-Burman. Herein lies the debate among linguists, anthropologists, and historians, over which population contained the initial inhabitants of the Indian region; some point to the Austric group, others believe the Dravidian group to have been in India first, and particular researchers think both language families stemmed from a common proto-Australoid language (Roychoudhury 2001). All sides can agree that the Tibeto-Burman group migrated from the Tibet and Myanmar areas (Roychoudhury 2001). An analysis of the mtDNA contrasted with language families provides results that suggest each language group in India represent a different founder group; however, the Austro-Asiatic group appears to have been the first to migrate to the Indian region (Roychoudhury 2001).

Previous studies in Asia found a north to south clinal graduation in genetic variation. A study by Suo et al. sought to understand the basis for this pattern. In examining SNPs from 22 populations, they found a strong correlation between allele frequencies and geographical latitude (Suo et al. 2012). A study by Qian et al. focuses on adaptations visible from genome-wide 63 populations equally representing linguistic and ethnic groups of Asia and supports the concept of local genetic adaptations. The researchers identified a selection for genes involved in hair follicle development and cancer. Southeast Asians demonstrated selection for genes involved in body mass, insulin, and metabolism regulation (Qian et al. 2013).

These modern scientific attempts to explain and oftentimes categorize human diversity have yielded varying interpretations. The lack of a consensus from examining linguistic and genetic patterns simultaneously makes an important point for the discussion of identity that will follow in the next chapter. Differing patterns of diversity can be interpreted to signify equally different historical accounts of a population. Therefore, whether linguistic and genetic patterns

reflect the same demographic events or not, such evidence does not dismiss cultural concepts of identity and heritage.

4 IDENTITY AND ANCESTRY TRACING

4.1 How Ancestry Tracing Impacts Identity

The increased availability of genetic technologies to the study of human populations has initiated a debate on both ethical identity and concerns. Commercial genetics services offer the unique opportunity for individuals to learn about their personal ancestral history (see Appendix D) whereas population geneticists, anthropologists, and epidemiologists consider the same topics of admixture, origins, and migrations but at the population or subpopulation levels. In contrast to researchers in academia that utilize all genetic markers depending on the research in question, commercial genetics most often utilizes haploid markers, specifically mitochondrial DNA or Y chromosome haplotypes, to make inferences on the ancestry of individuals, in lieu of autosomal markers. Since the mtDNA reflects only the maternal lineage and the Y chromosome reflects only the paternal lineage, this unilineal approach provides only half of a person's story. Additionally, the lineage approach can infer that two people or two groups share a common ancestor with relative confidence but it cannot infer exactly where this ancestral population would have resided and at what point in time they would have lived there. In other words, this method cannot soundly be applied for geographical inferences of the shared ancestral population. Furthermore, in instances where population data is unavailable some ancestry testing will incorrectly attribute portions of the genome to another population and skew the similarity between an individual's history and the samples available from that population. Origin inferences can also be incorrect when an individual represents a more recent admixture, and consequently, an origin that is rather intermediate between these two groups is attributed to be the origin. Before such information is released to the individuals and then the public, it is crucial

that extensive modeling is undertaken and all possible limiting factors are considered to restrict the amount of inaccurate inferences made on ancestral histories (Royal et al. 2010).

Our understanding of genetic ancestry does have the potential to be applied for the enhancement of population health. Since genetic factors oftentimes are the primary source for health risks, ancestry can be an important tool for gauging predisposition, especially in ethnic groups where particular genetic variations are exhibited more frequently. However, risk factors can be variable and may be attributable to environmental factors rather than genetic factors (Royal et al. 2010). In reality, each person displays a unique profile for risk factors, as every individual has a unique genetic makeup and distinctive environmental exposures based on their life choices (Risch et al. 2002). Regardless of the contributing factors, our understanding of genetic ancestry can serve as tool to educate and ensue behavioral changes for those that would have otherwise suffered hereditary diseases (Shrive and Kittles 2004). By categorizing potential risk factors, epidemiologists are able to put into effect plans for prevention and treatment, targeted to those that are most vulnerable. Ultimately, patients would be able to receive therapy that is even more individualized for their specific needs. Some researchers argue that the availability of genetic testing makes clinical evaluations based on genetics no longer an option but now an obligation for a good health practitioner (Risch et al. 2002).

The privacy of data collected for genetic ancestry tracing is a valid concern as this industry continues to grow. Since these commercial genetics endeavors are ultimately in the hands of corporations, should a corporation fail the future security of their data is of utmost interest. Even for companies that are still in business, there is justified anxiety caused by potential unauthorized sharing or using of individuals' personal genomic information. Despite the fact that they may never actually fall to the health risks preset by their genetic makeup, due to

chance, environmental factors, or behavioral choices, they could be penalized. For example, having a person's health genetic predispositions revealed to an insurance carrier could drastically impact the price they are required to pay for coverage or their ability to get coverage. (Royal et al. 2010).

Although the commercial genetics industry can provide a wealth of knowledge, the public is also gaining access to countless amounts of personal information in addition to inferences on population history being made that could be incorrect. While ancestry has been used interchangeably with "race" in the past, it is important to differentiate between these two terms and recognize ancestry as an origin associated with a particular parent population or geographic area. Societal implications resulting from the pursuit of popularizing genetic ancestry tracing have the potential to be quite severe. For groups in which membership is based on blood laws, genetic ancestry data could eliminate an individual from a group that they have identified with and lived within for their entire life (Royal et al. 2010, Elliott and Brodwin 2002). Obviously this is heavily based on the strictness of the group and how they perceive identity. Even in situations where individuals are still accepted by a group, they may suffer an identity crisis based on their results. While some may rely more heavily on etic perspective of how society classifies the ancestry of a person or a group based actual data, others may depend on a more emic view and base identity on the actual person or group's interpretation of their own identity or how they construct their personal genealogies. Nonetheless, new information regarding individuals' ancestry can have profound effects on how society perceives them and how they view themselves in the context of group and personal identities. (Royal et al. 2010). With these threats to identity, there is also room for a new basis of discrimination and injustice grounded on the abuse of genetic data. Ancestral categories determined by genetics could be

confused with racial categories, despite being recognized as attributes on a spectrum rather than with firm categorical boundaries (Shriver and Kittles 2004). More than on a personal level, such discrimination and injustice could surmount to the societal level. Cultural identities are upheld as banners of allegiance and alterations to such identities could have political implications. Identities and language are closely tied to nationalism and concepts of ethnicity. Having a shared bond of linguistic and cultural tradition imbues a sense of community and connectedness within a group of people (Rajagopalan 2001, Elliot and Brodwin 2002).

Concepts of identity are especially crucial in the discussion of the societal impacts of genetic ancestry tracing and are now an important point of investigation for anthropologists. Genetic tracing technologies are fulfilling much more than just the needs of academic research questions but now provide information to the public. Such technologies have the ability to instill a belief that our identities are born unto us and thereby unchangeable. More people are reevaluating who they think they are and who they can claim a social connection to. Embarking upon such studies poses a considerable threat to "personal esteem and self-worth, group cohesion, access to resources, and the redressing of historical injustice" (Brodwin 2002: 324). Quality control of interpretations extends beyond the genetic markers or sample size utilized in a study but also requires a comparison with other sources such as oral and cultural traditions and written records (Brodwin 2002).

Anthropologists have taken on this social issue for further investigation and pose questions to all aspects of the genetic ancestry tracing and identity conflict. Their investigation starts with an understanding of who is calling for these questions of ancestry, what audience is being provided the answers, historical timing at which people are asking these questions, and what occurring in the world or communities could be inspiring these questions into popular

subconscious (Brodwin 2002). Anthropologists seek to understand how genetic knowledge affects how people make claims to social groups and how they perceive themselves to be similar or dissimilar from someone else. Researchers also examine how the information provided by genetics can and should be incorporated into other sources of ancestry such as oral and archived evidence. Additionally, they are interested in why some groups place such value in genetic evidence and why others are highly skeptical (Brodwin 2002). Some classification is based on language family, geographic region, or political entity, while some groups are classified by multiple factors simultaneously. As the work of genetic ancestry tracing continues into the future, population studies may encourage the development of new ethnic groupings based on such data, a process known as ethnogenesis (Brodwin 2002).

A real danger is the possibility of genetic evidence evidentially undercutting other means of collective history, particularly if the public perceives science as the ultimate truth. In a time when the sources yield different results, a key question is which line of evidence should be accepted as most credible (Brodwin 2002, Elliot and Brodwin 2002). For example, in a situation where tracing is through the paternal or maternal line, a single ancestor that originates from an area distinct from the rest of the lineage and is of a more recent time period would shift the haploid typing and ultimately reflect a stronger connection to a different ancestry (Brodwin 2002). This scenario could easily discount a person's previously held identity, despite any other traditions or group membership that the individual and their family have upheld for generations. The classification of groups in genetic ancestry studies is also important for anticipating how the public will interpret any results (Brodwin 2002). Genetic information can strengthen or devalue previously held identities (Elliot and Brodwin 2002).

Shriver and Kittles (2004) emphasize the value of implementing clinical psychologists that are familiar with identity issues in the staff of a commercial genetic tracing group. By doing so, the company would be able to offer counseling to their customers and ensure their emotional and mental well-being after receiving their genetic results (Shriver and Kittles 2004). Additionally, this would remove the genetic essentialism mindset, or the idea that reduces humans to just a composite of their genetic makeup, and accentuates the human component of those being studied (Brodwin 2002). Furthermore, companies should be prepared to discuss the accuracy, application, and significance of ancestry tracing results with their customers (Shriver and Kittles 2004). Codes of conduct and accreditation have been proposed to uphold genetic tracing companies to a higher standard; yet, there is still too much concern for the liability of providing a stamp of approval for the inferences made from studies' results (Shriver and Kittles 2004, Royal et al. 2010). Anthropologists have the skills to serve as cultural brokers between the public and scientific communities; yet even the motives of such work are contested. If they are helping the public better understand the scientific perspective, should they not also help the scientists more fully understand the public's perspective? Experts must then reconcile whether they should aid in misunderstandings of genetic evidence if it means either advantageous or disadvantageous effects to a population. Whatever position experts in anthropology and other discipline experts choose to take, they cannot dismiss the weight of their opinion and how decisions can be made from their conclusions (Brodwin 2002).

4.2 Specific Cases

One famous example of a conflict founded on ancestry tracing, identity, and ownership is in the discovery of the Kennewick Man. The case begins with the unearthing of skeletal remains at the Columbia River near Kennewick, Washington in 1996. While some of his skeletal features made him appear to be a historic period Caucasoid. An archaic spear point lodged in his hip made this hypothesis seem incorrect. This conflicting evidence was assessed through Carbon 14 dating and ancient DNA analysis, but before a connection to either a Native American and Asian haplotype or a non-Native American and Asian haplotype could be established, the Army Corps of Engineers became involved. The Army Corps of Engineers managed the land where the remains were discovered and sought to repatriate the remains to the local Native American tribal group, per the Native American Grave Protection and Repatriation Act, known as NAGPRA (Kaestle and Smith 2005). This law

> Requires that the disposition of Native American remains discovered on federal lands or curated by federal agencies be determined by identifying their lineal descendants or 'cultural affiliation' with living Native Americans, if possible. Cultural affiliation is to be determined by 'a preponderance of the evidence based upon geographical, kinship, biological, archaeological, anthropological, linguistic, folkloric, oral traditional, historical or other relevant information or expert opinion' (NAGPRA, Section 7a(4)) (Kaestle and Smith 2005).

This inspired a protest from scientists that deemed such action to be a preemptive move, and consequently, the Burke Museum at the University of Washington detained the remains until ancestry could be determined by a scientific group chosen by the Department of the Interior. The group was unable to definitively determine any specific ancestry under nondestructive analyses; the skeletal morphology, however, most closely resembled modern Asian and Pacific Islander populations rather than either Native American or non-Native American populations. Even when scientists were permitted to use more invasive methods of inquiry, they were unable to find any

ancient DNA samples within the remains. A legal battle between the scientists and the local Native American tribes followed, ultimately granting scientific access to the remains once again (Kaestle and Smith 2005). In 2004, the Kennewick Man was determined to be most similar to the Moriori, a Polynesian group found near New Zealand but also sharing many features with the Ainu, possibly revealing a shared ancestral history (Walker and Owsley 2012).

Another famous case, which involved genetic ancestry testing revamping identity, is on the descendants of Sally Hemings and, supposedly, Thomas Jefferson. The study conducted Y chromosome haplotyping to examine this ancestral claim and pulled samples from descendants of Hemings, Jefferson's paternal uncle, and Jefferson's nephews' paternal grandfather. The claim was based on a historical account by Madison Hemings, one of Sally Hemings's children. In his account of his family history, he discusses how his mother was the daughter of a slave woman, Betty Hemings, and the plantation owner John Wayles, also Thomas Jefferson's father-in-law. When Wayles died, Betty and her children joined the plantation of Thomas and Martha Jefferson. When Sally Hemings was a teenager, she served as both a companion to Jefferson's daughter and as a mistress to Thomas Jefferson. She would bear six to seven children, one of which was relocated to the plantation of Thomas Jefferson's relative John Woodson after a scandalous story in the papers describing a slave child, known as "Tom" or "Thomas" that resembled Jefferson, and circulating rumors of his illegitimate children by a slave woman. From this new plantation family, Thomas and thereby his descendants took on the family name "Woodson" which is still carried by the Woodson family today. Genetic ancestry testing was pursued for this case and its results published in 1988; the data showed that one of Sally Hemings's other sons, Eston, had descendants whose Y haplotype precisely matched that of Thomas Jefferson's paternal grandfather but Thomas Woodson's descendants did not. In the

early 2000s, Sloan Williams (2005) examined the Woodson family's reactions to these results and the strength of the conclusions put forth. A major issue found was the lack of full disclosure in the consent form provided to the participants that provided samples; the consent form was clear on the procedures of the blood sampling process and with whom the ownership of the data would reside, but it was not explicit on why the study was being conducted or what possible consequences could arise. Unfortunately, despite an agreement to provide the results to the family prior to publication, most of the participants learned of the results for the first time from media sources or when the media contacted them for commentary. Additionally, the researchers from this study did not consider the detrimental impacts of revealing nonpaternity of a person to the rest of the Woodson family through this experiment. This case truly emphasizes the value of establishing trust and accountability with study participants (Williams 2005).

5 METHODS

5.1 The Population

The Lemba people are renowned in the research community for their claim to Jewish ancestry. The Lemba people live in parts of South Africa and Zimbabwe, and while many of their customs are similar to those in Jewish tradition, some researchers do not find this to be sufficient to confirm their Jewish ancestry. Genetic technologies have been implemented in population studies of the Lemba to scrutinize the validity of their claim. To understand how the Lemba might fit into the Jewish ancestry, it is valuable to first appreciate the Jewish history as a whole (Bjarnadottir 2013).

Jewish history is particularly remarkable because their cultural traditions have preserved a lineage that can be traced to tribes living in the Middle East around the second millennium BC (Atzmon et al. 2010). The present-day global Jewish population consists of roughly thirteen million people and can be divided into three main groups: Ashkenazi Jews, Sephardic Jews, and Middle-Eastern Jews. An extraordinary aspect of these populations is the shared conservation of their old traditions and customs, despite their initial separation some millennia ago. The Ashkenazi Jewish population consists of communities that migrated to Central and Eastern Europe. The Sephardic Jews are in communities that would have migrated to the Mediterranean and North Africa. The Middle Eastern Jewish population includes communities that remained in what were once the Babylonian and Palestinian regions. However, a large majority of Jewish people lives in either Israel or the United States (Hammer et al. 2000, Ostrer 2001).

The Jewish tradition asserts that their population originated with Abraham, as is described in the Old Testament of the Bible, at which point they were known as the Hebrews. Abraham, his son Issac, and his grandson Jacob would father all the descendants who would

suffer enslavement in Egypt and be eventually led to the "Promised Land." In time, this land and its inhabiting population would constitute the nation of Israel, representing both a religious and political entity (Zoloth 2003). Three major migration events from Israel would ultimately disperse its descendants across Eurasia, culminating in what is known as the Jewish Diaspora. In 586 BC, many Jewish people were forced to move to Babylon, in present day Iraq, after a temple in Jerusalem was demolished by Nebuchadnezzar (Diamond 1993). In 334 BC, Alexander the Great relocated numerous Jewish people to Egypt, Syria, and the Balkans. In 70 AD, many remaining Jewish people in Israel moved into the parts of Eurasia that were under the control the Roman Empire after the temple in Jerusalem was destroyed once again but by the Romans. By 300 AD, Jewish subpopulations founded new settlements in the Middle East, the Mediterranean, Europe, North Africa, and West Asia, specifically in present day Iraq, Germany, Spain, Italy, France, and Yemen, where they have resided up to the present day (Sachs and Bat-Miriam 1957).

With marriage traditions that inform marrying practices, it is possible that modern-day Jewish people are true descendants of the ancient population. However, Judaism represents both a religious and ethnic community, and there are some people who have converted to Judaism that would add an influx of new diversity and thereby dismantle this hypothesis. If the Lemba people live according to Jewish customs, a lineage back to the ancient Jewish population that modern Jewish groups claim could rightfully be theirs as well (Bjarnadottir 2013).

A major theory of interest to those that study Jewish ancestry is that of the Ten Lost Tribes of Israel. This theory suggests that ten tribes were exiled from Israel in the 6th century BC and were never heard from again. Claims of connections to these tribes have appeared in myriad parts of the globe, but it is unclear whether any modern group is actually related to any of the tribes or if the ten tribes of Israel actually existed and then disappeared. Historians are

particularly skeptical of some groups whose accounts are highly interwoven with missionaries and colonialism, as many methods of identification were forms established by Europeans (Parfitt and Semi 2002:53-64).

Genetic analysis has been used to evaluate claims of Jewish ancestry by many groups, including the Bantu-speaking Lemba people in South Africa. While Jewish people are not phenotypically homogenous, they do seem to be more similar within their population than compared to other non-Jewish populations. The genetic distinction present between the three Jewish groups matched the historical account of shared origin and then separation and isolation from each of the other groups. However, there was evidence of admixture with geographically neighboring populations, with the Jewish communities in Iraq and Iran demonstrating the least admixture (Ostrer and Skorecki 2013, Atzmon et al. 2010).

The Bantu and Khoisan populations have been of considerable interest to researchers from a variety of disciplines, particularly because they are recognized to be descendants of the deepest clades from which humans arose. Compared to genome-wide studies of other human populations, they are the most genetically divergent (Schuster et al. 2010). Perspectives on African history have changed over the last century as more archaeological work has been conducted in this region. Most of the initial work was based in colonial territories, and the theory of diffusion was supported as the basis for innovation and change in African societies. Researchers of this period argued that African populations had to have been influenced by Europeans, or some other non-African but equally civilized society. The concept of diffusion provided a valid explanation for how technology and culture was shared but was heavily loaded with racism. It was largely based on a perceived inability of Africans to create completely new technology and customs and then travel ad interact in such a way that would promote dispersion

of these new ideas. Free from a racist paradigm, diffusion remains an effective model for demonstrating the spread of culture. However, this perspective attributes archaeological remains in Africa to the local populations that resided there extensively, rather than to unknown and supposedly superior immigrants (Chami 2007).

Due to the age of Bantu societies, cultural, genetic, and archaeological evidence can often be compared to reveal patterns from history. In an investigation of settlement patterns in South Bantu groups, the size of the settlement, the extent of the territory, and the community's wealth was found to directly correlate with the levels in the political stratification. Early settlements established a hierarchy of power through ownership of cattle and met economic success mainly through trade of ivory (Huffman 1986). An interdisciplinary examination of Bantu people in South Africa has identified a correlation between linguistic and archaeological evidence based on geographic distribution and time period. Previous linguistic studies have attempted to use classification of languages to reconstruct relationships between extant and proto-languages (de Luna 2012). Settlement spatial distribution can be reconstructed using the point with the most diversity being set as the proto-language. The longevity of a proto-language is deemed to be directly correlated to the success of the settlements that speak it. Different aspects of language changed subtly over time, and its audience expanded as the original settlements grew and come into contact with other settlements. The correlation in linguistics with archaeology was found in ceramics from South Africa. De Luna (2012) hypothesizes that ceramic styles were highly connected to social identity and thus lends themselves for direct comparison with language. Both linguistic and archaeological evidence support a rapid and widespread expansion and divergence of founding Bantu populations. During a time when many

other settlements they encountered were hunter-gatherers, the Bantu practiced agriculture and iron working which greatly aided their swift development (Hiernaux 1968).

The Lemba people of South Africa, who consist of approximately 50-70,000 people divided into twelve clans, carry with them a strong oral and cultural tradition that claims Jewish ancestry and states that they came to Africa by boat (le Roux 2003). Their population lives among other groups, mainly in the Limpopo and Mpumalanga areas of South Africa and in southern Zimbabwe. Their connection to ancient Jewish populations is heavily disputed, by the scientific community but also within their group. Some group members are certain of this ancestral connection while others identify as being derived from Arabic origins (Soodyall 2013). Their linguistic culture seems to be their tie to the geographic area that they live in. Since many of the twelve clans live intermingled with Bantu-speaking groups, it is no surprise that many Lemba speak Bantu or other languages that are unique to the spatial range where they reside (le Roux 2003). In the oral tradition of the Lemba people residing in Zimbabwe, they claim northern origins in an area known as Sena, possibly in present day Yemen, Egypt, Judea, or Ethiopia, where they performed metalworking for the Arabs. The oral tradition of those in South Africa tells that their community came across the sea from Sena for trading purposes and eventually founded settlements. An elder of the Lemba Cultural Association provides another account that the Lemba had a community in Sena until Jerusalem was destroyed by the Babylonians, and exiles of this community emigrated. Eventually, the Lemba migrated in two waves to Ethiopia and by boat to South Africa and would launch trading routes along the eastern coast of Africa (le Roux 2003). Although there are some discrepancies among these oral accounts, an overall perspective can be derived: the Lemba believe their ancestry originates from outside of Africa and with the ancient Jewish tribes (le Roux 2003).

The Cohanim priesthood is one section of Jewish ancestry that is notably useful for investigating the Lemba's claim. According to biblical texts, this order dates back 3,300 years and can only be passed from father to son. This lineage is often denoted by some form of the surname "Cohen" and is believed to be directly descended from the brother of Moses, Aaron. These rules are extremely strict, and consequently the Cohen lineage includes only about 5% of the entire modern Jewish population (Skorecki et al. 1997, Jobling et al. 2004: 373-400, Bolster et al. 1998, Johnston 2003). A unique ancestral haplotype, Cohen modal haplotype (CMH), from the Y chromosome has been found to parallel the lineage of the Cohanim priesthood, being present in 60% of the males from the Cohen line, 12% in other Jewish males, and nearly nonexistent in non-Jewish males (Jobling 2001, Thomas et al 2000). In a study by Thomas et al, the Lemba Y chromosome was compared with Y chromosome samples from the Bantu in Africa, two types of Yemeni Jews, Ashkenazi Jews, and Sephardic Jews. The outcomes found the CMH haplotype on the Y chromosome of the Lemba men but not in the neighboring Bantu group; these were in accordance with the oral tradition of the Lemba with a Middle Eastern origin of either Arabic or Jewish descent. Furthermore, the Buba clan of the Lemba, which is regarded to be the oldest and highest ranking, demonstrated a frequency well over 50% for the CMH haplotype (Thomas et al 2000).

Most marriages are endogamous per Lemba customs, and male converts are usually not accepted into the tribe. This is less strict for women, who can be converted through a ritual much like the Jewish bath (Parfitt 2003). This has also been reflected in the mtDNA lineage, from which there seems to be no connection to Jewish ancestry. From this, it is possible to infer that Lemba men began to take wives from local African tribes upon their arrival and since that time (Thomas et al 2000). Their identification as Jews can be largely attributed to missionaries and

other travelers that found Lemba practices, such as male circumcision and a pork-free diet, to resemble that of noted Jewish populations (Parfitt 2003).

While genetic diversity studies pose considerable worth to participants, the impact of the media's coverage of the Lemba people's involvement in genetic ancestry tracing has been devaluing their identity. Their aims for recognition have been warped, and they have been racialized in South Africa based on their religion. However, their goal for the last seventy years has been to obtain state recognition as a distinct African ethnic group rather than being seen as Jews. Due to their small population size and other requirements to be considered a tribe, they were practically overlooked in the categorization and recognition by the state of South Africa of different groups residing in the area. During the apartheid and post-apartheid periods in South Africa, the Lemba were labeled as "black." Consequently, the Lemba have faced a tremendous amount of conflict between how society defines them racially and religiously, especially under the assumption that to be a Jew means to be white (Tamarkin 2011). The 1980s and 1990s brought an end to this political system and reshaped the way many people lived. The diverse number of groups classified as black South Africans were subjected to notably harsh limitations on geographic mobility; under this strict structure, people were required to have permits for both where they worked and where they lived. While they still face many barriers in society, they no longer need worry about being arrested or fined for their mobility (Reed 2013).

Much of the media has portrayed the Lemba identity through an essentialist perspective: if they possessed Jewish genetic traits, they must be Jews and always have been. This and other etic perspectives have clustered the Lemba with other groups, rather than prescribing the emic perspective and recognizing them as their own group. This lack of recognition has obstructed their day-to-day lives. Since being Lemban is not recognized as an ethnic identity by the state,

the Lemba people have been forced to assume an alternate ethnic identity in order to be attributed to a specific homeland region and in the passbooks they are required to carry. The Lemba people have a strong desire to distinguish themselves from being called one of the lost tribes of Israel as it makes them something that is exclusive to the past. Instead, they aspire to be recognized as a distinct ethnic group within South Africa, complete with the privileges that are provided with such acknowledgment (Tamarkin 2011).

5.2 Methods

This analysis of population studies will consult research utilizing genetics and linguistics to consider the value of pairing them as methods of demographic investigation. In my project, I consider the following questions, which I also describe in my introductory chapter:

1. Can one use linguistic groups as evolutionary units and is there enough evidence from population studies in genetics, geography, and linguistics to suggest that any of the three could be utilized to support the results of new studies from any of the others' types of data?
2. Specifically to the Lemba, are both linguistic and genetic evidence in support of their claimed Jewish heritage? What are the foreseen enhancements/consequences of this affirmation/contradiction to their self-prescribed identity?
3. How does modern policy need to be shaped to handle human rights in the human genome age? How can this case population be utilized to enhance ethics surrounding population ancestry tracing and identity?

I analyze genetic and linguistic data from open source databases, such as *Multitree* (2009) and Lewis et al. (2014), and previously published allele data. In scrutinizing population data and

previous studies, this analysis seeks to identify any and all correlations evident among genetic, linguistic, and geographic groups, and to identify regions or groups where these correlations are always or never present.

Alu insertion polymorphism frequencies are a component of the larger project dataset. These elements are either present or absent at a specific chromosomal location and are particularly advantageous in studying diversity in humans for several reasons. The presence of these polymorphisms is a derived characteristic, and therefore, descent is known with firm certainty when individuals or group possess the same Alu elements. Additionally, the rarity at which Alu elements are deleted makes it notably stable and useful for tracing lineages. p12F2 and 49a are restriction length fragment polymorphisms from the Y chromosome and are utilized as probes with the restriction enzyme TaqI (Mitchell 1996). The STRs are particularly useful for inferring demography due to their high mutation rate (Soodyall 2013).

Population classifications were taken strictly from the original data source. Alu insertion frequencies were available for individuals from Johannesburg, Richtersveld, and from the following populations: South African European, South African Jewish, Nama, Lemba, Venda, Zulu, Xhosa, Ndebele, Swazi, Southern Sotho, Pedi, Tswana, Tsonga, Ambo, Hetero, and Himba (see Appendix B). p12F2 frequencies were available for individuals from the following populations: Lemba, South African Jewish, South African European, South African Indian, Bantu, Khoisan, French, Czechoslovakian, Sephardic Jewish, Ashkenazi Jewish, Yemenite Jewish, Ethiopian Jewish, and Lebanese (see Appendix C). 49a frequencies were available for individuals from the following populations: Lemba, Negroid , Indian, European, South African Jewish, Sephardic Jewish, Ashkenazi Jewish, Falasha, and Yemenite Jewish (see Appendix A). STR data was available from the following populations: Lemba, Remba, Venda, and South

African Jewish (see Appendix E). The Remba are members of the Lemba population that split off and live in Zimbabwe and, unlike the Lemba, claim Arabic ancestry more than Semitic ancestry. The Venda are another group that lives in South Africa and, oftentimes, near the Lemba (Soodyall 2013). The terminology of these groups was maintained from the authors of the original data source, despite the problematic and debatable classification employed in deriving these categories. Language associations for each population were established using classifications from Comrie (1987), *Multitree* (2009), and Lewis et al. (2014).

6 RESULTS

6.1 Alu Insertion Data

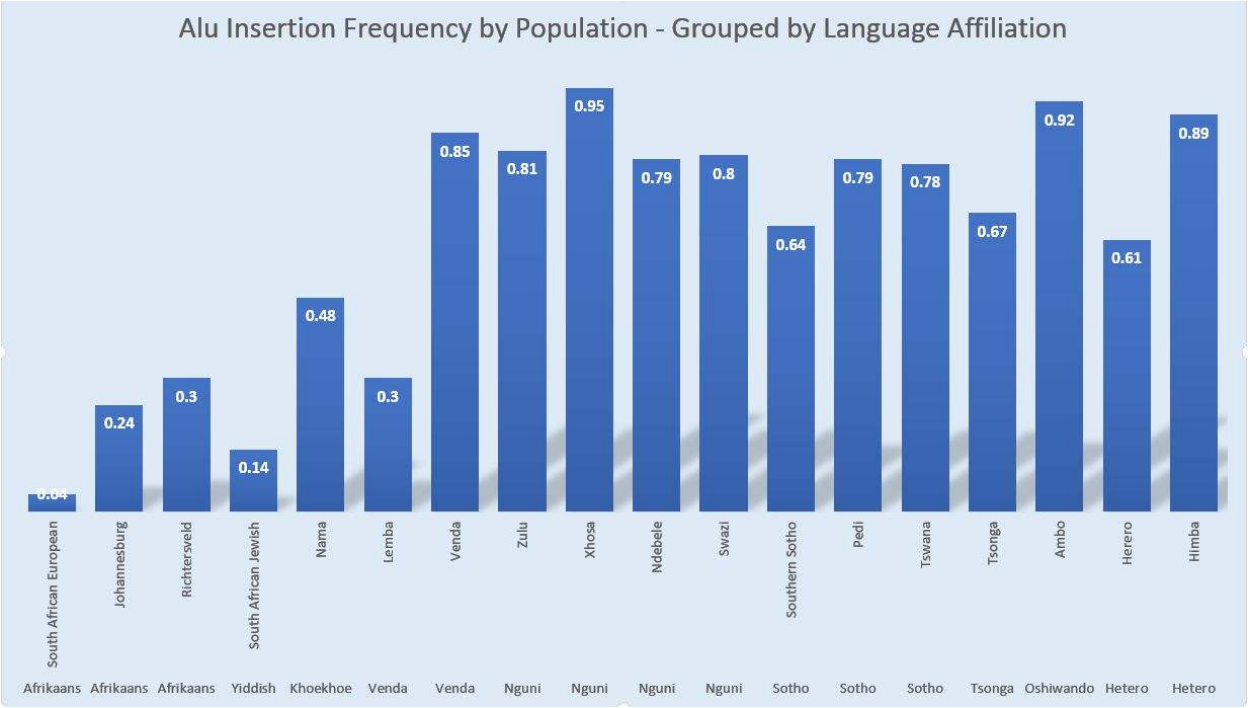

Figure 5, Alu Insertion Frequencies

The Lemba have an Alu Insertion frequency of 30%, which is more similar to the frequencies of the groups with mixed European ancestry, including South African European (4%), South African Jewish (14%), and individuals from Johannesburg (24%) and Richtersveld (30%). The other groups of Sub-Saharan Africa exhibit particularly high Alu Insertion frequencies, ranging from 61% and 95%. The next closest to the Lemba in level of frequency is the Nama group, with a frequency of 48%.

There does seem to be an association between Alu Insertion frequency and language affiliation. The members of the Nguni language community, which is also a part of the Bantu language group, demonstrate the highest Alu Insertion frequencies. The Alu Insertion

frequencies for members of the Sotho language community reflect a small range within the group. Similarly, the lowest frequencies are found both in the groups with the most European influence and who also are most affiliated with Afrikaans. Some interesting deviations from this pattern include the wider range present between the two groups from the Hetero language community as well as between the Lemba and Venda.

6.2 p12F2 Data

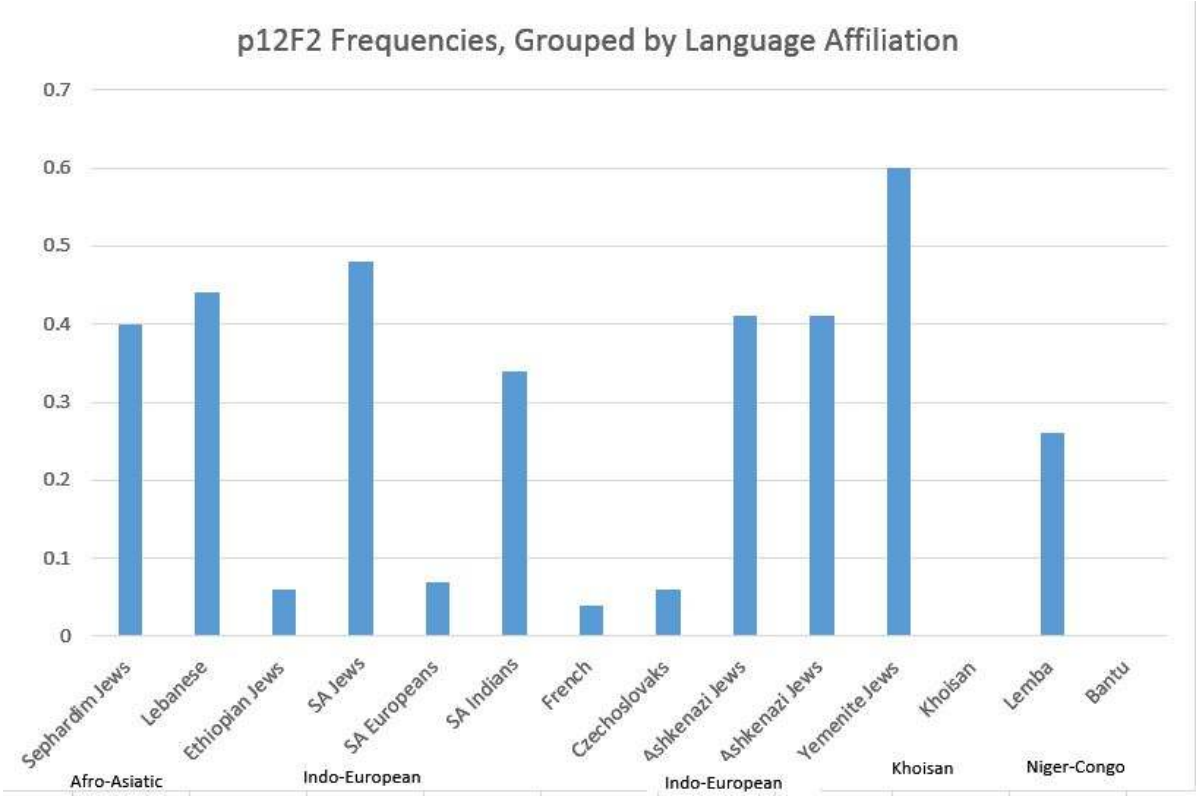

Figure 6, p12F2 Frequencies

The p12F2 polymorphism was not present in the individuals from the Bantu and Khoisan populations. It was visible in the lowest frequencies in the South African Europeans (7%), the French (4%), the Czechoslovakians (6%), and the Ethiopian Jewish (6%) populations. The Jewish populations, including Ashkenazi Jewish, Sephardic Jewish, and South African Jewish, demonstrated frequencies between 40-48% while the Yemenite Jewish population had a frequency of 60%. The Lemba and South African Indians had frequencies of 26% and 34%, respectively. In consideration of language affiliation with this data, these frequencies show no connection between language community and allele frequency. In this case, the Lemba frequencies more closely resemble those found in Jewish and Lebanese populations than those in the Bantu and Khoisan populations, of which they hold linguistic characteristics in common.

6.3 49a Data

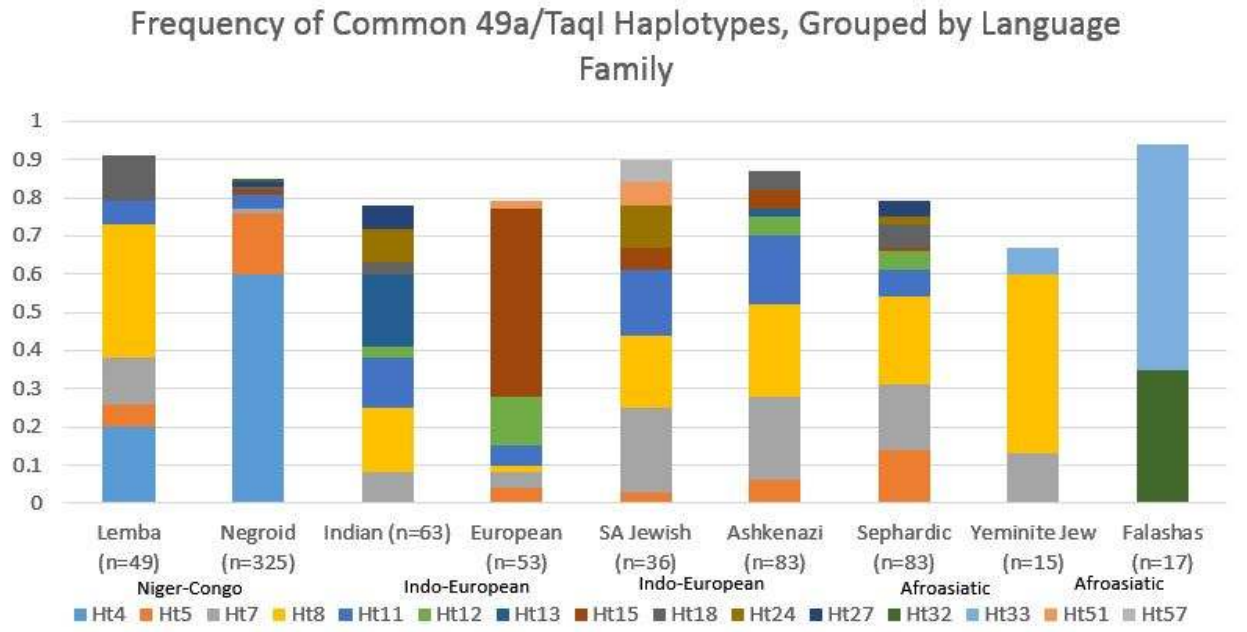

Figure 7, 49a Frequencies

The haplotypes Ht8 and Ht4 were most frequent of the haplotypes for the Lemba population. The Ht4 haplotype was also highly frequent while the Ht8 haplotype was absent in the Negroid population. The South African Jewish, Ashkenazi Jewish, and Sephardic Jewish populations had large frequencies of the Ht8 haplotype while no trace of the Ht4 haplotype. The Yemenite Jewish population possesses both the Ht4 and Ht8 haplotypes. The Lemba also had lower frequencies of the Ht5, Ht7, Ht11, and Ht18 haplotypes. The Ht5 haplotype appears in roughly 14-16% of the Negroid, as the group is defined by Spurdle and Jenkins 1996, and Sephardic Jewish populations whereas in only 3-6% in the Lemba, European, Ashkenazi Jewish, and South African Jewish populations. The Ht5 haplotype is not present at all in the Indian, Yemenite Jewish, and Falasha populations. The Negroid, Indian, and European populations have a range of 1-8% whereas the four Jewish populations have a range of 13-22% of the Ht7 haplotype. The Lemba have a 12%

frequency for the Ht7 haplotype putting it slightly in the middle of these two frequency ranges. The South African Jewish and Ashkenazi Jewish populations demonstrated higher frequencies for the Ht11 haplotype, at 17% and 18% respectively, compared to the Lemba (6%), Negroid (4%), European (5%), and Sephardic Jewish (7%) populations. These frequencies contrasted with language affiliation of these populations produces a similar result as is seen with the p12F2 frequency data and language communities and is dissimilar to the pattern found with the Alu insertion frequencies and language communities. In the Lemba population, for example, they are mainly affiliated with the languages of their geographically close neighbors, such as Bantu, whereas the 49a frequencies reflect a highest frequency of 35% from the Ht8 haplotype, much like the other Jewish populations. It is clear that the Lemba also share similar haplotypes with other Sub-Saharan populations.

6.4 6-STR haplotype Data

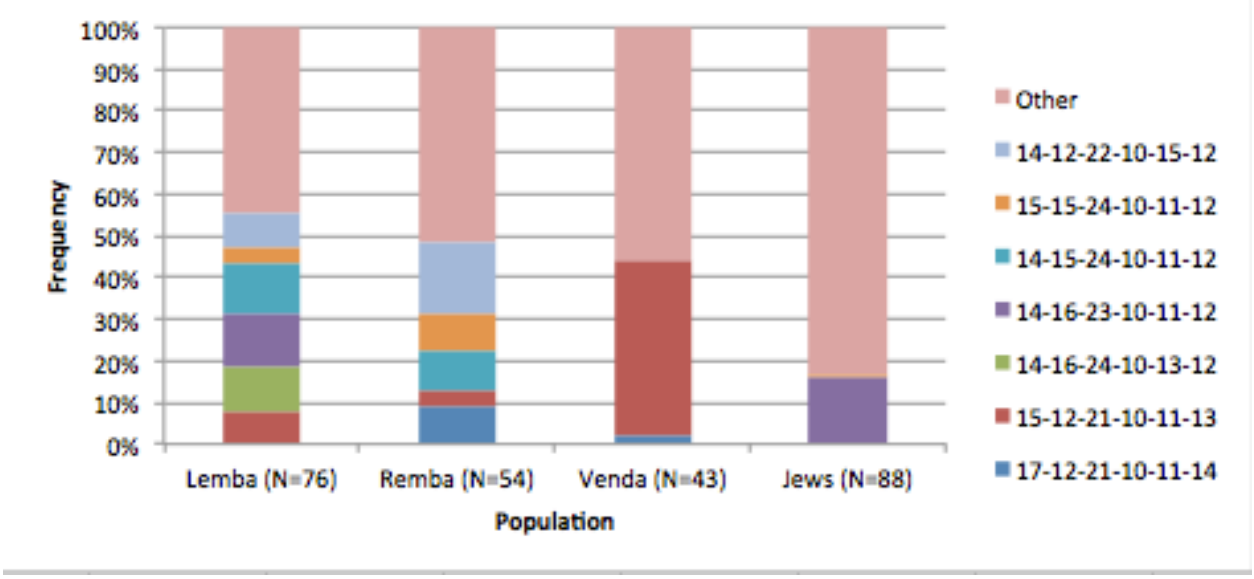

Figure 8, STR haplotype frequencies.

The 6-STR haplotype data compare haplotypes found in the Lemba population to those found in the Remba, Venda, and South African Jewish populations. The Lemba demonstrate haplotypes that appear in each of the other populations. The Remba have the most haplotypes in common with the Lemba, which is anticipated due to their known historically recent relationship. The haplogroup J, which is most closely affiliated with the Middle East (Soodyall 2013), is the most common found in the Lemba.

7 DISCUSSION AND CONCLUSION

7.1 Discussion

The three sources of genetic evidence considered yielded mixed results when compared with language communities of the populations in consideration. While there was some parallel in the Alu insertion frequencies and language, the frequencies of p12F2 and 49a compared with language suggested that linguistics and genetics were independent, despite being impacted by some of the same evolutionary forces. For each source, the Lemba demonstrated similar frequencies to the Jewish populations considered. However, as is visible in Figure 7, it is clear that haplotypes present in the Lemba are shared by both Sub-Saharan African and Jewish populations, suggesting influence from both.

The Alu insertion frequency in the Lemba was most similar to other populations that had been exposed to more recent Indo-European linguistic influence. The frequency of the p12F2 for the Lemba was particularly closer to the frequency of the Jewish populations, compared to the Bantu and Khoisan populations for which it did not appear at all. The frequency seen in the Lebanese population provides a counterpoint to a possible Lemba-Jewish relationship, as the Lebanese are not Jewish and exhibit similarly high frequency much like the Jewish populations. The frequency of the 49a haplotypes in the Lemba population demonstrated how the group could have mixed with different populations. The Ht4 haplotype that is found in high frequency in the Negroid population appears at roughly 20% in the Lemba population. The haplotype that appears most frequently in the Lemba, Ht8, is also found in similar frequency levels in the Jewish populations and not at all in the Negroid population. These results of similar frequencies in Lemba and Jewish and Lebanese populations require further investigation into both Semitic and

Arabic populations as possible influences. Additionally, the results do not seem to support the use of genetics and linguistics as predictors for the other.

The STR data is more recent and can be said to be a more state of the art comparison of the Lemba population to African and Jewish populations. According to Soodyall (2013), all four populations were deemed to be significantly different from each other based on mean pairwise differences and haplotype diversity. The finer resolution of Soodyall (2013)'s data does not support the Lemba's claim to Jewish ancestry. Cohanim modal haplotype (CMH) has previously been attempted as a justification for the support of the Lemban claim; yet, this data does not reflect the presence of the extended CMH in the Lemba or Remba populations. Arabic influence on the Lemba population is deemed to be more plausible than Semitic influence (Soodyall 2013).

Despite the contradictions from the genetic evidence, the Lemba have held a claim to Jewish ancestry for some time; their assertion became more vocal during the rise of apartheid in South Africa, at a time when ethnic identity was so important. The establishment of the Lemba Cultural Association has promoted the awareness of Lemban identity conflict. The organization appeared in the 1940s as a reaction to strong European influence and a desire to distinguish themselves from other African communities. Many blacks of South Africa were alienated from their land to permit the European intrusion of the property. For this and cultural reasons, the Lemba people were becoming more dispersed throughout South Africa without regularly reconvening and maintaining communication, and consequently, the culture and oral history was at risk of being lost. Work by ethnologists and missionaries of the nineteenth century, particularly Henri Junod, attempted to construct a taxonomy of languages and establish distinct tribes in South Africa based on this taxonomy and cultural customs. The Lemba did not have a language that could be completely separated on the basis of individuality and therefore were

consequently lumped in with other tribal groups. While they were noted for their cultural similarities to Semitic traditions even at this time, they adopted language from the African communities that they resided near. The movement to increase awareness of the Lemba identity has been met with resistance. The Lemba have also been considerably discriminated against by other African groups, particularly the Venda, and such prejudices have been evident in their ability to get and hold jobs. Many people, especially Lembans in college, do not acknowledge their identity amongst their peers. In fact, the people most removed from urban areas and progressed in age will openly claim the Lemban identity mainly (Bujis 1998).

South Africa is particularly diverse in its languages spoken, as the home to the major language families of Khoisan, Niger-Congo, Indo-European, and Sign Language. Bantu languages, which are a part of the Niger-Congo family, represent the largest group of speakers in this region (Mesthrie 2002: 11-25). Previous studies have demonstrated possible correlations existing between genetic and linguistic evidence, oftentimes using populations such as the Bantu as a model to prove such point. From such a conclusion, linguistic family trees have been used to support phylogenetic trees (Cavalli-Sforza et al. 1988). This study with the Lemba clearly demonstrates that such a parallel between these two sources of evidence is not always the case. While population size could be a factor impacting these different results, this could also be a reflection of more recent linguistic changes for the Lemba population.

Ideologies of ethnicity kept the prosperity of South Africa unanimously with the white minority prior to 1994. Since that time, wealth has been distributed across a hierarchy with Anglophone whites at the top and then the Afrikaners, Indians, coloureds, and Africans. For much of the twentieth century, many people were not openly racist but placed emphasis on biological and cultural differences, often attempting to rationalize segregation through the

influence of cultural anthropology. The sharp boundaries between African ethnic communities perpetuated the white minority's ethnocentrism and secured their position of power in South Africa by conquering through division of the majority. More recently, there has been a greater recognition for the sub-African ethnicity. While separation by ethnic group is still utilized in more rural areas, Africans living in more urban settings do not claim previously held ethnic labels or homelands (Glaser 2001: 132-160).

The Lemba Cultural Association has sought to inspire Lemban people to own their heritage by promoting Lembans that have achieve great success. They hope to perpetuate the achievements of their population by hosting meetings about careers, talents, and skills, especially for children. Additionally, the LCA holds an annual conference to re-instill the oral tradition and cultural values of their heritage and clearly illustrates what makes them distinct from other groups residing in South Africa (Bujis 1998). Bujis (1998) suggests that the strong push to be recognized as a distinct ethnic group can be largely attributed to their struggle for jobs and resources (Bujis 1998).

7.2 Conclusion

The genetic evidence produced supporting the Lemba claim to Jewish ancestry plays an important role in this project's discussion of identity. Researchers are forced to ask how much value genetics adds when it is used to confirm or disprove a previously held identity or social connection. Establishing knowledge of ancestry is inevitably going to change a population's future. While ancestry is a part of the Lemba's past, it will fundamentally shape the connections that they make in the world in the present and how they perceive themselves and those outside of their group. For the Lemba, the genetic evidence was a tangible confirmation of the identity that

they had hoped to seek recognition for; yet, this validation begs us to consider if an identity based on cultural tradition is not sufficient to stand alone (Brodwin 2002). In other words, do results from genetic tracing hold more significance than oral or written tradition? Genetic evidence also has the potential to complicate the ways in which groups deem individuals to be worthy of membership. When originally one could be discerned for membership based on his parents or other visible factors, genetic inheritance could bring new ways to rank the legitimacy of individuals' claims to an ancestry. This possibility could profoundly alter methods of ethnic inclusion and exclusion and thereby restrict rights and obligations to a certain few (Brodwin 2002).

This project's recognition of the impact of population studies on human rights and our ability as scientists to communicate with the public could be fundamental in inhibiting the creation of new forms of racialization. With new methods of distinguishing individuals from one another, it is inevitable that both differences and similarities will be revealed but it is possible that either could receive too much attention. An additional effort in this study is to address the potential of a resurged eugenics movement during the human genome era. The eugenics movement of the early 20th century is easily scolded today as extreme and cruel; however, during its time, it was both socially acceptable and a component of the mainstream life (Kelves 2011). What is dismissed as pseudoscience when looking retrospectively could be the accepted science of the present. For this reason, it is crucial to include this perspective in this study's investigation of population studies and identity (Kelves 2011).

Physical anthropologists have both perpetuated the evidence for and aided in the fight against concepts of physical race based on human variation. In the last century, racist attitudes have shifted considerably but are still present in many parts of the world. Where "races" or

"subspecies" are no longer studied, many groups termed "populations" are still considered through a racial lens (Caspari 2003). The discussion of eugenics is also applicable to a debate in the quality of life in modern biomedicine. With prenatal testing permitting detection of disabilities at very early stage in life, debates surround issues such as who in this situation would really know what is "best" and who has the freedom to act on these choices (Miller and Levine 2013).

Ultimately, increased research in population studies directly translates to better understanding of the impacts of social and cultural factors on a population and has implications for the disciplines of biomedicine, forensics, and anthropology (Bisol et al. 2008). Both linguistic and genetic studies can benefit from an enhanced perspective on identity and the role of scientific research in shaping identity. It is an exciting endeavor into what makes us similar, what makes us different, and what makes us human. However, this path must be taken with care; as we gain new knowledge, we are responsible for any consequences that erupt from this wisdom. Categorized groups of any kind are susceptible to forms of genetic profiling. It is this project's goal to fully understand the methodology behind the research in demography studies and serve as a cultural broker between the scientific and public perspectives, unlike what has occurred in the cases of Kennewick Man, the Sally Hemings and Thomas Jefferson families, and the Lemba population (Brodwin 2002).

REFERENCES

2009. Multitree: A digital library of language relationships. Ypsilanti, MI: Institute for Language Information and Technology (LINGUIST List), Eastern Michigan University. http://multitree.org/

Genetic kit for ancestry. 23andMe. https://www.23andme.com/

DNA Ancestry. Ancestry. http://dna.ancestry.com/

Atzmon, Gil, Li Hao, Itsik Pe'er, Christopher Velez, Alexander Pearlman, Pier Francesco Palamara, Bernice Morrow, Eitan Friedman, Carole Oddoux, Edward Burns, and Harry Ostrer. 2010 Abraham's Children in the Genome Era: Major Jewish Diaspora Population Comprise Distinct Genetic Clusters with Shared Middle Eastern Ancestry. *American Journal of Human Genetics*. 86(6): 850-859.

Barbujani, Guido, and Robert R Sokal.
1990 Zones of sharp genetic change in Europe are also linguistic boundaries. *Proc Natl Acad Sci USA* 94:4516-4519.

Barbujani, Guido.
1991 What do languages tell us about human microevolution? *Trends Ecol. Evol.* 6:151-155.

Barbujani, Guido, and Andrea Pilastro.
1993 Genetic evidence on origin and dispersal of human populations speaking languages of the Nostratic macrofamily. *Proc. Natl. Acad. Sci. USA* 99: 4670-4673.

Barbujani, Guido, Gerard N. Whitehead, Giorgio Bertorelle, and Ivane S. Nasidze.
1994 Testing Hypotheses on Processes of Genetic and Linguistic Change in the Caucasus. *Human Biology* 66(5): 843-864.

Barbujani, Guido.
1997 DNA Variation and Language Affinities. *Am. J. Hum. Genet*. 61:1011-1014.

Belle, Elise M.S., and Guido Barbujani.
2007 Worldwide Analysis of Multiple Microsatellites: Language Diversity has a Detectable Influence on DNA Diversity. *American Journal of Physical Anthropology*. 133:1137-1146.

Bertorelle G, and Guido Barbujani.
1995 Analysis of DNA diversity by spatial autocorrelation. *Genetics* 140:811-819.

Bisol, Giovanni Destro, Paolo Anagnostou, Chiara Batini, Cinzia Battaggia, Stefania Bertoncini, Alesso Boattini, Laura Caciagli, Carla M. Calo, Cristian Capelli, Marco Capocasa, Loredana Castri, Graziella Ciani, Valentina Coia, Laura Corrias, Federica Crivellaro, M. Elena Ghiani,

Donata Luiselli, Cristina Mela, Alessandra Melis, Valeria Montano, Giorgio Paoli, Emanuele Sanna, Fabrizio Rufo, Marco Sazzini, Luca Taglioli, Sergio Tofanelli, Antonella Useli, Giuseppe Vona, and Davide Pettener.
2008 Italian isolates today: geographic and linguistic factors shaping human biodiversity. *Journal of Anthropological Sciences*, 86:179-188.

Bjarnadottir, Sigrun.
2013 On the Jewish ancestry of the Lemba people of South Africa. *How genetics can provide information to support historical claims of ancestry.*

Boas, Franz.
1912 Changes in Bodily Form of Descendants of Immigrants. *American Anthropologist*. 14(3): 530-562.

Bolster, JS, RR Hudson, and SJ Gaulin.
1998 High Paternal Certainties of Jewish Priests. *American Anthropologist*. 100(4): 967-971.

Brodwin, Paul.
2002 "Genetics, Identity, and the Anthropology of Essentialism." *Anthropological Quarterly* 75(2): 323-330.

Brown, Steven, Patrick E. Savage, Albert Min-Shan Ko, Mark Stoneking, Ying-Chin Ko, Jun-Hun Loo, and Jean A. Trejaut.
2013 Correlations in the population structure of music, genes, and language. *Proc. R. Soc. B.* 281:1-7.

Buijs, G.
1998 Black Jews in the Northern Province: A study of ethnic identity in South Africa. Ethnic and Racial Studies. 21:661-682.

Caspari, Rachel.
2003 From Types to Populations: A Century of Race, Physical Anthropology, and the American Anthropological Association. *American Anthropologist*. 105(1):65-76.

Carmelli, D. and L.L. Cavalli-Sforza
1979 The genetic origin of the Jews: a multivariate approach. *Hum. Biol*. 51:41-61.

Casanova, M., P Leroy, C Boucekkine, J Weissenbach, C Bishop, M Fellous, M Purrello, G Fiori, and M Siniscalco.
1985 A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science*. 230(4732): 1403-1406.

Cavalli-Sforza, Luigi-Luca.
2000 *Genes, Peoples, and Languages*. North Point Press: New York, NY.

Cavalli-Sforza LL, A Piazza, P Menozzi, and J Mountain.

1988 Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proc Natl Acad Sci USA* 85:6002-6006.

Chami, Felix A.
2007 Diffusion in the Studies of the African Past: Reflections from New Archaeological Findings. *African Archaeological Review*. 24:1-14.

Colonna, Vincenza, Alessio Boattini, Cristina Guardiano, Irene Dall'Ara, Davide Pettener, Giuseppe Longobardi, and Guido Barbujani.
2010 Long-Range Comparison between Genes and Languages Based on Syntactic Distances. *Human Heredity* 70: 245- 254.

Comrie, Bernard (ed.).
1987 The World's Major Languages. Oxford University Press.

De Luna, Kathryn M.
2012 Surveying the Boundaries of Historical Linguistics and Archaeology: Early Settlement in South Central Africa. *African Archaeological Review*. 29:209-251.

Diamond, J.
1993 Who are the Jews? *Natural History*. 102(11): 12-19.

Elliot, Carl, and Paul Brodwin.
2002 Identity and genetic ancestry tracing. *BMJ* 325(7378): 1469-1471.

Genetic Testing for Ancestry, Family History, and Genealogy. FamilyTreeDNA. https://www.familytreedna.com/

Genographic Project. National Geographic. https://genographic.nationalgeographic.com/

Glaser, Daryl.
2001 Politics and Society in South Africa. SAGE Publications. London.

Gonzalez-Galarza, FF., S. Christmas, D. Middleton, and AR Jones.
2011 Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. Nucleic Acid Research. 39:D913-D919.

Gravel, Simon, Fouad Zakharia, Andres Moreno-Estrada, Jake K. Byrnes, Marina Muzzio, Juan L. Rodriguez-Flores, Eimear E. Kenny, Christopher R. Gignoux, Brian K. Maples, Wilfried Guiblet, Julie Dutil, Marc Via, Karla Sandoval, Gabriel Bedoya, The 1000 Genomes Project, Taras K. Oleksyk, Andres-Ruiz Linares, Esteban G. Burchard, Juan Carlos Martinez-Cruzado, and Carlos D. Bustamante.
2013 Reconstructing Native American Migrations from Whole-Genome and Whole-Exome Data. *PLoS Genetics*. 9(12): e1004023.

Gringer.
2013 Simulation of common example used describing the effect random sampling has in genetic drift. Web image. Accessed November 24, 2014. < http://en.wikipedia.org/wiki/Genetic_drift#mediaviewer/File:Random_sampling_genetic_drift.svg>.

Hammer, Michael F., AJ Redd, ET Wood, MR Bonner, H Jarjanazi, T Karafet, S Santachiara-Benerecetti, A Oppenheim, MA Jobling, T Jenkins, H Ostrer, and B Bonne-Tamir.
2000 Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. *PNAS USA*. 97(12): 6769-6774.

Harding, Rosalind M., and Robert R. Sokal.
1988 Classification of the European language families by genetic distance. *Proc. Natl. Acad. Sci. USA* 85: 9370-9372.

Hiernaux, Jean.
1968 Bantu Expansion: The Evidence from Physical Anthropology Confronted with Linguistic and Archaeological Evidence. *The Journal of African History*. 9(4): 505-515.

Huffman, Thomas N.
1986 Archaeological Evidence and Conventional Explanations of Southern Bantu Settlement Patterns. *Africa: Journal of the International African Institute*. 56(3): 280-298.

Jobling, Mark A.
2001 In the name of the father: surnames and genetics. *Trends in Genetics*. 7(6): 353-357.

Jobling, Mark A., Mathew Hurles, and Chris Tyler-Smith.
2004 Human Evolutionary Genetics: Origins, Peoples and Disease. *Garland Science*. 1st edition.

Johnston, J.
2003 Case study: the Lemba. *Developing World Bioethics*. 3(2): 109-111.

Kaestle, Frederika A., and David G. Smith.
2005 Working with ancient DNA: NAGPRA, Kennewick Man, and Other Ancient Peoples. In Turner, Trudy R., ed. 2005 Biological Anthropology and Ethics: From Repatriation to Genetic Identity. Albany, NY: *State University of New York Press*, p. 241-262.

Kelves, Daniel L.
2011 From Eugenics to Patents: Genetics, Law, and Human Rights. *Annals of Human Genetics*. 75(3): 326-333.

Kroeber, A.L.
1926 *Anthropology: Biology and Race*. Harcourt, Brace, and World, Inc. New York, NY.

Lee, S. and T. Hasegawa

2014 “Oceanic barriers promote language diversification in Japanese islands.” *Journal of Evolutionary Biology*.

Le Roux, Magdel.
2013 The Lemba – a lost tribe of Israel in Southern Africa? *Unisa press*.

Lewis, Paul M., Gary F. Simons, and Charles D. Fennig (eds.).
2014 *Ethnologue: Languages of the World*, 17th Edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com

Marginalia, Professor.
2009 Representation of a population bottleneck. Web image. Accessed November 24, 2014. < http://en.wikipedia.org/wiki/Genetic_drift#mediaviewer/File:Population_bottleneck.jpg>.

Marginalia, Professor.
2009 Representation of the founder effect. Web image. Accessed November 24, 2014. < http://en.wikipedia.org/wiki/Genetic_drift#mediaviewer/File:Founder_effect_with_drift.jpg>.

Mesthrie, Rajend.
2002 “South Africa: a sociolinguistic overview” in Mesthrie, R., ed. 2002 Language in South Africa. Cambridge University Press.

Molnar, Stephen.
1983 *Human Variation: Races, Types, and Ethnic Groups*. Prentice-Hall, Inc. Englewood Cliffs, NJ.

Monsalve, MV., A. Helgason, and DV Devine.
1999 Languages, geography, and HLA haplotypes in Native American and Asian populations. *Proc. R. Soc. Lond. B*. 266: 2209-2216.

Montagu, Ashley.
1962 The Concept of Race. *American Anthropologist*. 64(5): 919-928.

Mielke, James H., Lyle W. Konigsberg, and John H. Relethford.
2011 Human Biological Variation, 2nd Edition. Oxford University Press.

Miller, Paul Steven, and Rebecca Leah Levine.
2013 Avoiding genetic genocide: understanding good intentions and eugenics in the complex dialogue between the medical and disability communities. *Genetics in Medicine*. 15:95-102.

Mitchell, RJ.
1996 Y-Chromosome-Specific Restriction Fragment Length Polymorphisms (RFLPs): Relevance to Human Evolution and Human Variation. *American Journal of Human Biology*. 8:573-586.

Nasidze, Ivane, and Mark Stoneking.

2001 Mitochondrial DNA variation and language replacements in the Caucasus. *Proc. R. Soc. Lond. B*. 268:1197-1206.

National Human Genome Research Institute.
2003 Different Variation of Phenotype Due to Environment. Web image, accessed November 24, 2014. < http://geneed.nlm.nih.gov/topic_subtopic.php?tid=48>.

Nei, Masatoshi and Gregory Livshits.
1989 Genetic Relationships of European, Asians, and Africans and the Origin of Modern *Homo Sapiens*. *Hum Heredity*. 39:276-281.

Nettle, Daniel, and Louise Harriss.
2003 Genetic and Linguistic Affinities between Human Populations in Eurasia and West Africa. *Human Biology* 75(3): 331-344.

Nichols, Johanna.
1997 "Modeling Ancient Population Structures and Movement in Linguistics." *Annual Review of Anthropology* 26:359-384.

Ostrer, Harry.
2001 A genetic profile of contemporary Jewish populations. *Native Reviews Genetics*. 2(11): 891-898.

Ostrer, Harry, and Karl Skorecki.
2013 The population genetics of the Jewish people. *Human Genetics*. 132(2): 119-127.

Parfitt, Tudor, and Emanuela Semi.
2002 Judaising Movements: Studies in the Margins of Judaism in Modern Times. *Routledge*.

Parfitt, Tudor.
2003 Constructing black Jews: genetic tests and the Lemba – the 'black Jews' of South Africa. *Developing World Bioethics*. 3(2): 112-118.

Poloni, ES, O Semino, G Passarino, AS Santachiara-Benerecetti, I Dupanloup, A Langaney , and L Excoffier.
1997 Human genetic affinities for Y chromosome p49a,f/*Taq*I haplotypes show strong correspondence with linguistics. *Am J Hum Genet* 61:1015-1035.

Qian W, L Deng, D Lu, and S Xu.
2013 Genome-wide Landscapes of Human Local Adaptation in Asia. *PLoS ONE* 8(1): e54224.

Rajagopalan, Kanavillil.
2001 The Politics of Language and the Concept of Linguistic Identity. *CAUCE* 24:17-28.

Reed, Holly E.

2013 Moving Across Boundaries: Migrations in South Africa, 1950-2000. *Demography*. 50:71-95.

Risch, Neil, Esteban Burchard, Elad Ziv, and Hua Tang.
2002 Categorization of humans in biomedical research: genes, race, and disease. *Genome Biology*, 3(7): 1-12.

Ritte, U, E Neufeld, M Broit, D Shavit, and U Motro.
1993 The differences among Jewish communities- maternal and paternal contributions. *Journal of Molecular Evolution*. 37:435-440.

Rosenberg, Noah A., Jonathan K. Pritchard, James L. Weber, Howard M. Cann, Kenneth K. Kidd, Lev A. Zhivotovsky, and Marcus W. Feldman.
2002 Genetic structure of Human populations. *Science*. 298(5602): 2381-2385.

Royal, Charmaine D., John Novembre, Stephanie M. Fullerton, David B. Goldstein, Jeffrey C. Long, Michael J. Bamshad, and Andrew G. Clark.
2010 Inferring Genetic Ancestry: Opportunities, Challenges, and Implications. *The American Journal of Human Genetics*, 86:661-673.

Roychoudhury, Susanta, Sangita Roy, Analabha Basu, Rajat Banerjee, H. Vishwanathan, M.V. Usha Rani, Samir K. Sil, Mitashree Mitra, and Partha P. Majumder. 2001 Genomic structures and population histories of linguistically distinct tribal groups of India. *Hum Genet*., 109:339-350.

Sachs, Leo, and Mariassa Bat-Miriam.
1957 The Genetics of Jewish Populations. 1. Finger Print Patterns in Jewish Populations in Israel. *American Journal of Human Genetics*. 9(2): 117-126.

Santachiara Benerecetti, AS, O Semino, G Passarino, A Torroni, R Brdicka, M Fellous, and G Modiano.
1993 The common, Near-Eastern origin of Ashkenazi and Sephardi Jews supported by Y-chromosome similarity. *Annual of Human Genetics*. 57:55-64.

Scheinfeldt, Laura B, Sameer Soi, and Sarah A. Tishkoff.
2010 Working toward a synthesis of archaeological, linguistic, and genetic data for inferring African population history. *PNAS* 107(2): 8931-8938.

Schuster, Stephan C., Webb Miller, Aakrosh Ratan, Lynn P. Tomsho, Belinda Giardine, Lindsay R. Kasson, Robert S. Harris, Desiree C. Petersen, Fangqing Zhao, Ji Qi, Can Alkan, Jeffrey M. Kidd, Yazhou Sun, Daniela I. Drautz, Pascal Bouffard, Donna M. Muzny, Jeffrey G. Reid, Lynne V. Nazareth, Qingyu Wang, Richard Burhans, Cathy Riemer, Nicola E. Wittekindt, Priya Moorjani, Elizabeth A. Tindall, Charles G. Danko, Wee Siang Teo, Anne M. Buboltz, Zhenhai Zhang, Qianyi Ma, Arno Oosthuysen, Abraham W. Steenkamp, Hermann Oostuisen, Philippus Venter, John Gajewski, Yu Zhang, B. Franklin Pugh, Kateryna D. Makova, Anton Nekrutenko, Elaine R. Mardis, Nick Patterson, Tom H. Pringle, Francesa Chiaromonte, James C. Mullikin,

Evan E. Eichler, Ross C. Hardison, Richard A. Gibbs, Timonthy T. Harkins, and Vanessa M. Hayes.
2010 Complete Khoisan and Bantu genomes from southern Africa. *Nature*. 463:943-947.

Shriver, Mark D., and Rick A. Kittles.
2004 Genetic ancestry and the search for personalized genetic histories. *Genetics*, 5:611-618.

Skorecki, Kari, Sara Selig, Shraga Blazer, Robert Bradman, Neil Bradman, PJ Waburton, Monica Ismajlowicz, and Michael F. Hammer.
1997 Y-chromosomes of Jewish priests. *Nature*. 385(32)

Soodyall, H.
2013 Lemba origins revisited: Tracing the ancestry of Y-chromosomes in South Africa and Zimbabwean Lemba. *S. Afr. Med. J*., 103(12 Suppl 1):1009-1013.

Spurdle, Amanda B. and Trefor Jenkins.
1996 The Origins of the Lemba "Black Jews" of Southern Africa: Evidence from p12F2 and other Y-chromosome markers. *American Journal of Human Genetics*. 59: 1126-1133.

Spurdle, Amanda B., Michael F. Hammer, and Trefor Jenkins.
1994 The Y Alu Polymorphism in Southern African Populations and Its Relationship to Other Y-specific polymorphisms. *American Journal of Human Genetics*. 54:319-330.

Suo, Chen, Haiyan Xu, Chiea-Chuen Khor, Rick TH Ong, Xueling Sim, Jieming Chen, Wan-Ting Tay, Kar-Seng Sim, Yi-Xin Zeng, Xuejun Zhang, Jianjun Lu, E-Shyong Tai, Tien-Yin Wong, Kee-Seng Chia, and Yik-Ying Teo.
2012 Natural positive selection and north-south genetic diversity in East Asia. *European Journal of Human Genetics*. 20:102-110.

Tamarkin, Noah.
2011 Religion as Race, Recognition as Democracy: Lemba "Black Jews" in South Africa. *The Annals of the American Academy of Political and Social Science*, 637:148-164.

Thomas, Mark G., Tudor Parfitt, Deborah A. Weiss, Karl Skorecki, James F. Wilson, Magdel le Roux, Neil Bradman, and David B. Goldstein.
2000 Y chromosomes traveling south: the Cohen modal haplotype and the origins of the Lemba-the "Black Jews of Southern Africa". *American Journal of Human Genetics*. 66(2):674-686.

Veeramah, Krishna R., and Michael F. Hammer.
2014 The impact of whole-genome sequencing on the reconstruction of human population history. *Nature Reviews: Genetics*. (15):149-162.

Ward, RH, A Redd, D Valencia, B Frazier, and S Paabo.
1993 Genetic and linguistic differentiation in the Americas. *Proc Natl Acad Sci USA* 90: 10663-10667.

Williams, Sloan R.
2005 A Case Study of Ethical Issues in Genetic Research: The Sally Hemings-Thomas Jefferson Story. In Turner, Trudy R., ed. 2005 Biological Anthropology and Ethics: From Repatriation to Genetic Identity. Albany, NY: *State University of New York Press*, p. 185-208.

Zoloth, L.
2003 Yearning for the long lost home: the Lemba and Jewish narrative of genetic return. *Developing World Bioethics*. 3(2):128-132.

APPENDICES

Appendix A – 49a/TaqI Frequency Data

Appendix A.1

Haplotype	L	N	I	E	SA Jews	A Jews	S Jews	Y Jews	F
Ht4	0.2	0.6	0	0	0	0	0	0	0
Ht5	0.06	0.16	0	0.04	0.03	0.06	0.14	0	0
Ht7	0.12	0.01	0.08	0.04	0.22	0.22	0.17	0.13	0
Ht8	0.35	0	0.17	0.02	0.19	0.24	0.23	0.47	0
Ht11	0.06	0.04	0.13	0.05	0.17	0.18	0.07	0	0
Ht12	0	0	0.03	0.13	0	0.05	0.05	0	0
Ht13	0	0	0.19	0	0	0.02	0	0	0
Ht15	0	0.01	0	0.49	0.06	0.05	0.01	0	0
Ht18	0.12	0.01	0.03	0	0	0.05	0.06	0	0
Ht24	0	0	0.09	0	0.11	0	0.02	0	0
Ht27	0	0.01	0.06	0	0	0	0.04	0	0
Ht32	0	0.01	0	0	0	0	0	0	0.35
Ht33	0	0	0	0	0	0	0	0.07	0.59
Ht51	0	0	0	0.02	0.06	0	0	0	0
Ht57	0	0	0	0	0.06	0	0	0	0

Table 1, 49a/TaqI Haplotype Frequencies (Data from Spurdle and Jenkins 1996).

Appendix A.2

Abbreviation	Population	Sample Size
L	Lemba	49
N	Negroid	325
I	Indian	63
E	European	53
SA Jews	South African Jews	36
A Jews	Ashkenazi Jews	83
S Jews	Sephardic Jews	83
Y Jews	Yemenite Jews	15
F	Falashas	17

Table 2, Sample Populations (Data from Spurdle and Jenkins 1996).

Appendix B – Alu Insertion Frequency Data

Population				**Observed Freq.**	**SE**	**N**	**Observed Values**
Caucasoid							
	South African European			0.04	0.03	51	2.04
	South African Asiatic Indian			0	0	63	0
	South African Jewish			0.14	0.06	36	5.04
	Total			0.05	0.02	150	7.5
Mixed Ancestry							
	Johannesburg			0.24	0.05	66	15.84
	Richtersveld			0.3	0.08	33	9.9
	Total			0.26	0.04	99	25.74
Khoisan							
	Nama			0.48	0.1	23	11.04
	Tsumkwe San			0.11	0.05	38	4.18
	Sekele San			0.44	0.07	45	19.8
	Total			0.46	0.06	68	31.28
Enigmatic Bantu-speakers							
	Lemba			0.3	0.07	47	14.1
Negroid							
	Eastern Bantu						
		Nguni					
			Zulu	0.81	0.06	47	38.07
			Xhosa	0.95	0.05	22	20.9
			Ndebele	0.79	0.11	14	11.06
			Swazi	0.8	0.07	30	24
	Sotho/Tswana						
		Southern Sotho		0.64	0.07	45	28.8
		Pedi		0.79	0.06	52	41.08
		Tswana		0.78	0.07	40	31.2
	Tsonga			0.67	0.09	30	20.1
	Venda			0.85	0.07	27	22.95
	Western Bantu						
		Ambo		0.92	0.04	38	34.96

		Herero					
			Herero	0.61	0.07	46	28.06
			Himba	0.89	0.05	35	31.15
			Total	0.78	0.02	426	332.28
Khoisan-speaking Negroid							
	Dama			0.77	0.07	37	28.49
Pygmy							
	Pygmy			0.79	0.08	24	18.96

Table 3, Alu Insertion Frequencies (Data from Spurdle and Jenkins 1994).

Appendix C – p12F2/TaqI Frequency Data

Population (n)	Frequency (Standard Error)
Lemba (46)	.26 (0.6)
SA Jews (29)	.48 (.09)
SA Europeans (43)	.07 (.04)
SA Indians (53)	.34 (.07)
Bantu (182)	.00 (.00)
Khoisan (90)	.00 (.00)
French (26)	.04 (.04)
Czechoslovaks (100)	.06 (.02)
Sephardim Jews (80)	.40 (.06)
Ashkenazi Jews (80)	.41 (.06)
Lebanese (88)	.44 (.05)
Ashkenazi Jews (44)	.41 (.07)
Yemenite Jews (15)	.60 (.13)
Ethiopian Jews (17)	.06 (.09)

Table 4, p12F2/TaqI Frequencies (Data from Spurdle and Jenkins 1996, Casanova et al 1985, Ritte et al. 1993, Santachiara Benerecetti et al 1993).

Appendix D – Genetic Ancestry Companies

Service Provider	Type of DNA considered	Type of Results
DNA Ancestry	Autosomal, 700,000 markers	Recent genetic ethnicity, combines with historical records
23andme.com	Autosomal, Y chromosome, mtDNA	Ancestry, no longer offer Health, comparison to Neandertals
The Genographic Project	Autosomal, Y chromosome, mtDNA	Ancestry, migrations
FamilyTreeDNA	Autosomal, Y chromosome, mtDNA	Ancestry, reconstructing genealogies

Table 5, Genetic Ancestry Companies

Appendix E – 6-STR Haplotype Data

Haplogroup	Haplotype	Lemba (N=76)	Remba (N=54)	Venda (N=43)	Jews (N=88)
E-M2	17-12-21-10-11-14		0.093	0.023	
E-M2	15-12-21-10-11-13	0.079	0.037	0.419	
J-12f2a	14-16-24-10-13-12	0.105			
J-12f2a*	14-16-23-10-11-12	0.132			0.159
J-12f2a	14-15-24-10-11-12	0.118	0.093		
J-M172	15-15-24-10-11-12	0.039	0.093		0.011

L-M11	14-12-22-10-15-12	0.079	0.167		

Table 6, STR Frequency data from Soodyall (2013)