The History of Admixture in African Americans

JESSICA M. GROSS

University of New Mexico - Main Campus

Follow this and additional works at: https://digitalrepository.unm.edu/anth_etds

Part of the Anthropology Commons

Recommended Citation
https://digitalrepository.unm.edu/anth_etds/176

This Dissertation is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Anthropology ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.
Jessica Marie Gross

Candidate

Anthropology

Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Heather JH Edgar (chair)

Sherry Nelson

Yann Klimentidis

David Prior
Dedication

To Patrick Nagatani and Keith Hunley who inspired me to think creatively and encouraged me to trust and pursue my ideas.

To my parents, Jeanne Katsuko Crawford and Walter Gene Gross, who sparked my interest in the study of race and admixture.
Acknowledgements

I thank all of the people who agreed to participate in the studies that provided the data used in my dissertation. Although I did not collect these samples, nor were the data collected specifically for my studies, the willingness of the participants to engage with scientists allowed me to conduct this research. I also thank all of the scientists and museums who made the data available to me and other researchers.

My success in graduate school would not have been possible without the support of my advisor and committee chair, Heather JH Edgar. I am grateful to you, Heather, for your willingness to read everything I sent, for being readily available to meet with me to discuss my research, and for encouraging me to move forward and publish it. You were everything I could have asked for in an advisor, and, without you, this document would not exist. That alone would have been enough, but you have been much more than an advisor to me, you have been my friend and advocate, and for that, I am deeply grateful. I enjoyed all of our conversations about both anthropology and art, and I will miss having them regularly. Thank you so much for everything, Heather.

I am thankful to the members of my committee, Yann Klimentidis, Sherry Nelson, and David Prior. Thank you for sharing your thoughts on my research, for your willingness to meet with me to discuss it, and for your generosity in accommodating my schedule. I greatly appreciate your time and insight and am grateful for our conversations.
Patty Crown, signing up for your proposal writing class was one of the best decisions I made in graduate school. I am grateful to have learned how to clarify my ideas, write grant applications, and critique my peers’ work from someone with such a keen mind. You were much more than an instructor to me, you were a role model and a friendly, supportive presence throughout my time as a graduate student. It was always a welcome surprise to run into you on campus, and I will dearly miss seeing you regularly.

I am grateful to my parents, Jeanne Katsuko Crawford and Walter Gene Gross, who taught me by example that, through hard work and dedication, I can persevere even when the odds are against me. Thank you, mom, for supporting me financially and emotionally with kindness, grace, and a noticeable lack of judgement. Thank you, dad, your pride in me made me want to try harder and do better.

I thank Tracy Leonard, Donnie Leonard, and Donna Wiley for your kindness and generosity. For the first time in my life, I feel like I am part of a loving family. Keith Pallesen, thank you for always pushing me, inspiring me to be my best, and treating me like your own child.

Thank you to all of the staff in the Anthropology Department. Jennifer George, you have been kind and friendly and were always incredibly efficient and helpful. Thank you for making everything so easy. JoNella Vasquez, thank you for helping me with my finances and answering all of my questions, and for taking time to talk to me every time I saw you. Thank you, Matt Tuttle for always
greeting me with a smile. You made being in the department vastly more enjoyable, and I will miss stopping by your office for a conversation.

Thanks to everyone in the Bridge Lab and to Drew Enigk, Stephanie Fox, and Darragh Hare for listening to my presentations, giving me feedback, and sharing with me a friendly environment to talk about everything related to graduate school, research, and life. Thank you to Marieken Shaner and Dorothy Scholl for giving me funding, and making me feel valued.

Thank you to all of my friends. I leaned on each of you when graduate school and life got tough, and you picked me up with your kindness, support, and humor. Thank you, Meg Healy, Jenny Cabotage, and Bonnie Young, for making my transition to graduate school seamless by readily incorporating me into your lives. The combination of fun times (barbecues, ANTM, game nights, Halloween parties, etc.) and serious discussions about the woes of research made my graduate school experience feel complete and fulfilling. Kris Sabbi, Kate Rusk, and Marian Hamilton, I am so glad to have shared this experience with you. Although our days of half-priced wine are over, I will always have fond memories of spending my Mondays with you. Lexi O’Donnell, thank you for your unconditional friendship, for weekly lunches and coffee breaks, and for your eagerness to help with my research. I have greatly valued your friendship and camaraderie. Renee McKitterick, Jessamyn Lovell, Mia Sosa-Provencio, Lea Quale, and Dorielle Caimi, Carolyn Marsden, and Robin Nelson, each of you has shared with me your friendship and wisdom, you made me stronger and
encouraged me to trust my mind and advocate for myself. Thanks, Nicole Caimi, for reminding me that nothing really matters except kindness, friendship, and health. Thank you for our Friday ramen sessions, Beth Stelle. I think of our conversations when I am struggling to move forward. Thank you, Sam Wohletz, for repeatedly reminding me that I am an okay person when things were falling apart around me. Jazlyn Mooney and Sara Niedbalski, in addition to being wonderful colleagues, you have been true friends and were constant sources of positivity, which kept me going. James Kwan, I am so grateful that you remained my friend for all of these years despite my ridiculous schedule and reclusiveness. Our meetups were a welcome break from the stress of grad school and they helped me continue moving forward. Jim Stone, you helped shape my mind, and rightly alerted me to the fact that anthropology was just as much a dead end as art, so I may as well do what makes me happy. Brooke Steiger and Yoshiko Shimano, words cannot express my gratitude for the sanctuary you so generously gave me. I am forever in your debt. I am also so grateful for your friendship and support, Brooke. I would have lost my mind without you. Martin Muller, thank you for helping me. I am grateful for your clear and logical mind and for your friendship. Finally, with all my heart, thank you, Helen Davis and Cristin Deuble, for welcoming me with open arms to be your unofficial roommate, and for taking care of me ever since. It is the collective effort of all of you that carried me to this point. My goal upon entering graduate school was to gain an
understanding of humanness, and from all of you, I learned about the best parts of the human condition.

Thank you, Patrick Nagatani, my dear friend and mentor, for your endless positivity and encouragement. You inspired in me creativity and wonder, opened my eyes to the world, and showed me the joy that is being human. You taught me to see the beauty in humanity, to have a sense of humor in tough times, and to treasure every moment of life. I applied all of these lessons in graduate school, and could not have made it through without them. You were likely the best human I will ever know, and my life is vastly richer for having known you. I will never forget you, Patrick. Love and Namaste.

Lastly, but most importantly, I thank Keith Hunley who shared with me his wealth of knowledge of human evolution, and fostered my desire to understand humanness. I will treasure our many conversations about life, beauty, literature, music, and art, which added light to this experience. Thank you, dear friend, you’re the best anthropologist since Boas.
The History of Admixture in African Americans

By

Jessica Marie Gross

B.A. Anthropology & Art Studio, University of New Mexico, 2008
B.F.A. Art Studio, University of New Mexico, 2009
M.A. Anthropology, University of California, Riverside, 2010
M.S. Anthropology, University of New Mexico, 2013
Ph.D. Anthropology, University of New Mexico, 2019

ABSTRACT

African American admixture has been a focal topic of genetic studies since the mid-20th century. Generally, these studies estimate individual- and population-level African and European ancestry proportions. Some of these studies fit unrealistic admixture models to the patterns of genetic diversity in African Americans to determine both the onset time of admixture between Africans and Europeans, and the per-generation contribution of Europeans.
This research has failed to consider the contribution of the millions of Africans who migrated, either forcibly or by choice, to North America during and after slave importation, and failed to consider how changing social dynamics have affected the interactions between people of African and European descent over the past 400 years. These social dynamics include forcible control of mating by slave owners, antebellum and post-Civil-War segregation, the Great Migration, and the Civil Rights Movement. In this research, I demonstrate how this oversight has clouded our understanding of the admixture process in African Americans, including the timing and degree of African and European contributions, and prevented us from exploring the effects of social processes on the interactions between human populations across human evolution.

I attempted to overcome these shortcomings using historical, genetic, and dental data to estimate the timing and degree of European and African contributions to African Americans in different regions of the US. First, I fit two models of admixture to genetic datasets that included African contribution. For these models, I used historical data to inform the per-generation African contribution, then compared the fit of these models to the unrealistic models used previously. Second, after determining that dental morphological data can be used for ancestry estimation, I evaluated geographic and temporal variation in African American ancestry using dental morphological characteristics, and addressed the possible causes of that variation.
Important conclusions about African American admixture from each of the three studies that I conducted for this thesis are: 1) admixture models that incorporate history about African Americans fit empirical ancestry distributions better than models that fail to account for this history, 2) the variation in ancestry across time and space is due to admixture and possibly drift.

TABLE OF CONTENTS

ABSTRACT ........................................................................................................................................... ix
LIST OF FIGURES ................................................................................................................................. xiii
LIST OF TABLES ................................................................................................................................. xiv
CHAPTER 1. Introduction ..................................................................................................................... 1
  Studies of African American admixture ..................................................................................... 1
  Morphological studies of human variation ........................................................................... 5
  Guide to the dissertation ...................................................................................................... 7
CHAPTER 2. Tests of fit of historically-informed models of African American admixture .......................................................................................................................... 12
  Abstract ........................................................................................................................................ 12
  Introduction ............................................................................................................................. 14
  Methods .................................................................................................................................... 20
  Results ....................................................................................................................................... 30
  Discussion and conclusions .................................................................................................. 39
CHAPTER 3. Informativeness of dental morphology in ancestry estimation in African Americans ......................................................................................................................... 47
  Abstract ....................................................................................................................................... 47
  Introduction ............................................................................................................................. 48
  Materials and methods ........................................................................................................ 52
  Results ....................................................................................................................................... 58
  Discussion and conclusions .................................................................................................. 65
CHAPTER 4. Variation in African ancestry among African Americans across space and time ......................................................................................................................... 72
  Abstract ....................................................................................................................................... 72
  Introduction ............................................................................................................................. 74
Methods ...............................................................................................................................78
Results ....................................................................................................................................81
Discussion .............................................................................................................................86
Conclusions ..........................................................................................................................93
CHAPTER 5. Summary and Conclusions .............................................................................94
Appendix ..................................................................................................................................101
References .............................................................................................................................102
LIST OF FIGURES

Figure 2.1. African migration ................................................................. 18
Figure 2.2. Generalized admixture model .............................................. 24
Figure 2.3. Individual ancestry distributions ......................................... 32
Figure 2.4. Expected model distributions over time .................................. 34
Figure 2.5. ASW IA distributions for 12 generation models ....................... 37
Figure 2.6. AME IA distributions for 12 generation models ....................... 38
Figure 2.7. ABT IA distributions for 12 generation models ....................... 39

Figure 3.1. Fisher Information ............................................................... 59
Figure 3.2. STRUCTURE results for all datasets ...................................... 62
Figure 3.3. STRUCTURE output for reduced-FI genetic datasets .............. 65

Figure 4.1. Distributions of Cluster 1 membership ................................... 84
Figure 4.2. PCA 1 and 2 ..................................................................... 86
LIST OF TABLES

Table 2.1. Sample sizes and number of loci in the three African American samples ................................................................. 22
Table 2.2. Model parameters each generation .......................................................................................................................... 23
Table 2.3. Possible mating pairs after generation 1 in each model .............................................................................................. 25
Table 2.4. Mean and range of African American ancestry estimates for the ASW, AME, and ABT samples ........................................... 31
Table 2.5. Model-fitting results for the ASW, AME, and ABT samples at \( g = 12 \) generations ................................................................. 36

Table 3.1. Fisher Information .................................................................................................................................................. 60
Table 3.2. Cluster 1 membership ............................................................................................................................................. 64
Table 3.3. Correlation in Cluster 1 membership .......................................................................................................................... 65

Table 4.1. African American samples used in analyses ........................................................................................................... 79
Table 4.2. STRUCTURE results for Cluster 1 membership ....................................................................................................... 82
Table 4.3. Proportions of individuals who differ from sample mean African ancestry .................................................................................. 83
Table 4.4. Pairwise Wilcoxon Rank Sum Tests ............................................................................................................................. 85

Table A.1. Dental morphology data description ......................................................................................................................... 101
Table A.2. Genetic data description ........................................................................................................................................ 101
CHAPTER 1. INTRODUCTION

As an evolutionary anthropologist, I am interested in the ultimate causes of human variation, and in how social and cultural factors shape diversity within our species over time. My thesis focuses on admixture in African Americans, a population which comprises approximately 13% of US population (US Census Bureau, 2010). Admixture is the mixing of long separated populations via the exchange of mates. The African American population arose 12 generations ago in North America through mating between slaves taken forcibly taken from sub-Saharan Africa, primarily western Africa, and European slave owners, and variation in this population has been shaped by the continued contribution of Africans and Europeans since then. In evolutionary terms, this mixing is a special case of gene flow. Gene flow homogenizes populations, and maintains unity within our species, while notions of race separate us. The countervailing notions of unity and difference have characterized the history of the African American population and they have strongly influenced genetic research on African American admixture.

Studies of African American admixture

Since the mid-20th Century, anthropologists and geneticists have studied the admixture process in African Americans using two broad approaches. The first approach uses various statistical methods to estimate the proportion of African and European ancestry in African American populations and individuals.
This research has revealed that African American populations in all regions of the US have ancestry from both European and African “source” populations, and that African and European ancestry proportions vary among individuals within and among these regions (Glass and Li, 1953; Bryc et al., 2015; Baharian et al., 2016). The second approach estimates the timing and degree of European contributions to the African American population by fitting unrealistic admixture models to extant patterns of genetic diversity in African Americans (Glass and Li, 1953; Long, 1991; Gravel, 2012; Kidd et al., 2012; Bryc et al., 2015; Baharian et al., 2016).

In terms of ancestry estimates at the population and individual levels, the principal findings of these studies are that: 1) the mean African ancestry level in extant US African American samples is always substantially higher than the mean European level (Glass and Li, 1953; Long, 1991; Parra et al., 2001; Bryc et al., 2015), 2) the mean Native American ancestry level is relatively low in African Americans throughout the US (Parra et al., 1998; Tishkoff et al., 2009; Bryc et al., 2015), 3) despite high mean African ancestry proportions at the population level, African ancestry varies substantially among individuals, from a low of about 2% to a high of about 98% (Bryc et al., 2015), and 4) more mating has occurred between European males and African or African American females than between African American males and European females (Parra et al., 1998, 2001; Bryc et al., 2015).
In terms of the dynamics of the admixture process, studies have produced ambiguous results. Bryc et al. (2015), for example, fit a discrete two-stage admixture model to distributions of ancestry tract lengths and concluded that a single admixture event occurred between Africans and a combined European/Native American group six generations before the present. Baharian et al. (2016) found that a two-pulse admixture model, in which Europeans contribute to the African American population on two separate occasions, fits the data of extant African Americans better than a one-time admixture event model, and that the first incident of admixture occurred around 1740. Glass and Li (1953) assumed a model of continuous gene flow into the African American population from Europeans, and estimated a per-generation rate of European contribution to the African American population of 3.58%. More recently, Parra et al. (2001) concluded that a model of continuous one-way gene flow from Europeans with a per-generation gene flow rate of between 2.0 to 3.1% is consistent with the distribution of European alleles in African American individuals. Even more recently, Jin et al. (2011) fit four admixture models, including a continuous a two-way gene flow model, to the distribution of ancestry tracts for a sample of African Americans. The authors constrained their admixture onset time to between 10 and 17 20-year generations in the past in an effort to be consistent with African American history, and found that one-way continuous gene flow from Europeans into the African American population with an onset of 14 generations in the past produced the best-fit distribution of ancestry tracts.
The results of these studies are broadly incompatible with one another, and, for the most part, they are inconsistent with our current understanding of African American history. Remarkably, with the exception of those fit by Jin et al. (2012), this research has failed to consider the contribution of the millions of Africans who migrated forcibly or by choice to North America during and after slave importation, and it has failed to consider how changing social dynamics have affected the interactions between people of African and European descent over the past 400 years. Jin et al. (2011), however, did not ground the per-generation source-group contributions in history. Rather, they used observed ancestry fractions, and divided the contribution from each source population evenly across the generations in their model.

To my knowledge, no study of genetic admixture has yet incorporated ongoing African (mostly western, sub-Saharan African, but also from other regions of Africa) input into the African American population based on historical records, despite the fact that it may have been the most prevalent source of gene flow into the African American population for generations after the onset of slavery. Additionally, no previous study tested whether shifts in social convention and laws over the course of African American history had a significant impact on the current pattern of variation among African Americans.
Morphological studies of human variation

Another potential source for ancestry estimation is morphological data. An advantage of morphological data is that, analogous to ancient DNA, it permits us to study population movements and admixture in the past. In the case of African Americans for example, it permits us to directly examine changes in the timing and amount of admixture prior to and following key historical events. Morphological data are also inexpensive to obtain, and do not require specimen destruction.

There is a long history of using cranial morphological traits to reconstruct evolutionary relationships among populations (Berry and Berry, 1967; Ossenberg, 1976; Jose et al., 2001; Hanihara et al., 2012; Movsesian, 2013). These traits have been shown to be heritable, suggesting that variation in traits reflects underlying genetic diversity (Sjøvold, 1984; Velemínský and Dobšíková, 2005; Carson, 2006). Furthermore, analyses of cranial morphological traits have identified similar patterns of global and regional human diversity as neutral genetic data (Hanihara et al., 2003; Relethford, 2004; Roseman, 2004). The results of these studies indicate that cranial morphology is a potential proxy for neutral genetic diversity, and therefore appropriate for use in studies of evolution.

Similar work has been done using dental morphological data. Analyses of dental morphological data, for example, have identified a positive correlation between diversity and geographic distance (Irish and Guatelli-Steinberg, 2003), and shown that patterns of within and between regional variation are consistent
with a serial founder effect process (Hanihara, 2008). These results indicate that
dental morphological data are potentially suitable for use in studies of evolution.

The cranial and dental morphological studies mentioned above focus on
diversity on the global scale, and not on individuals in admixed populations such
as African Americans. In recent years, researchers have used morphological
data to estimate population affinities for individuals (Edgar, 2007, 2015; Irish,
2015). For example, Edgar (2007) calculated biodistance between samples of
African Americans and European Americans to approximate the change over
time in the European contribution to the African American population. In a more
recent study, Irish (2015) used presence/absence scores for dental traits to
assign ancestry to individuals of unknown origin. This study assigned a
consensus ancestry based on a simple majority of trait features, meaning a
single ancestry was assigned to each individual. Edgar (2015) showed that
groups differ at a multivariable level, and produced a discriminant function to
assign individuals to race groups.

Algee-Hewett (2016) used a cluster-based algorithm similar to a method
used in genetic studies to estimate ancestry from cranial morphological data and
compared her results to those published in genetic studies. The goal of this study
was to test whether ancestry could be reliably estimated using cranial
morphology without the use of source populations. She found that the results of
her analyses corresponded to those in the published studies in terms of patterns
of population structure and patterns of admixture.
Dental morphological data have been used in biological distance studies and to assign individuals to populations or race groups, but, to date, with one exception, they have not been used to estimate ancestry proportions within individuals. However, the results of these studies, along with the work of Algee-Hewett (2016) demonstrate that it may be possible to use dental morphological data to estimate individual-level ancestry and reconstruct the evolutionary process in African Americans, which was the goal of this research.

The exception is Delgado et al. (2018) who estimated individual European, Native American, and African ancestry in Colombians. However, Delgado et al. used only 16 traits to estimate ancestry, and based their interpretations on comparisons with ancestry estimates made from genetic data using a completely different method. Since it is unclear how the estimates relate, the accuracy of their analysis is unclear.

Guide to the dissertation

It is in this context that I conducted my dissertation research. My overarching goal was to better understand African American admixture given the changing dynamics of African American and European American relations over the course of US history as evidenced and influenced by laws, social conventions, and immigration. This dissertation is comprised of three data analysis chapters (Chapters 2, 3, and 4) written for peer review publications. The first two chapters (Chapters 2 and 3) have been published in The American
Journal of Physical Anthropology (Gross, 2018; Gross and Edgar, 2019), and the third chapter will be submitted this year. In all three chapters, I use the model-based clustering method implemented in the program STRUCTURE to estimate individual biogeographic ancestry. My focal population is African Americans and I use West Africans, and Europeans and European Americans as source populations for ancestry estimation. I use two genetic datasets and one dental morphological dataset for my analyses. The first genetic dataset consists of 992,601 single nucleotide polymorphisms genotyped in 271 West African, European, and African American individuals. The second genetic dataset consists of 645 short tandem repeat loci genotyped in 177 West African, European, and African American individuals. The dental morphological dataset consists of 62 traits scored in 797 West African, European, and African American individuals.

My goal in Chapter 2 was to improve our understanding of the admixture process in the African American population by modeling the continuous contribution of both Africans and Europeans to the African American population. This chapter is novel in that it is the first study to directly incorporate slave importation data, and it is the first chapter to model gene flow into African Americans from both Africans and Europeans. I estimated the individual African and European ancestry in the two genetic datasets and produced distributions of individual African ancestry in the African Americans. Next, I wrote an original program to run the mechanistic model devised by Verdu and Rosenberg (2011).
Parameters of the models are the per-generation contributions from source populations, and number of generations. I altered the parameters to run four different versions of the model: 1) one-time admixture, 2) one-way gene flow from Europeans, 3) two-way gene flow from Europeans and African Americans, and 4) two-way gene flow with ancestry-related assortative mating. In Models 3 and 4, I used information on African immigration to North America during and after slavery (Curtin, 1969; Rawley, 1970; Gibson and Lennon, 1999) to determine the per-generation rate of African contribution to the African American population. In model 4, I devised a resampling method to simulate ancestry related assortative mating, and wrote an original program in R to implement the simulation.

In Chapter 3, I determine the relative power of dental morphology in ancestry estimation, and to establish that a model-based clustering method frequently applied to genetic data could be applied to dental morphological data. A secondary goal of this study was to determine whether nondichotomized morphological traits perform better than dichotomized versions of the same traits. This chapter served as necessary preparation for Chapter 4, in which I examined geographic and temporal variation in ancestry in African Americans.

For the following analyses, I used both of the genetic datasets and the dental morphological dataset, and used the dental traits in both dichotomized and nondichotomized form. First, I calculated the Fisher Information of each dataset to determine whether the data had the potential to be informative about ancestry
in the first place. Next, I ran the program STRUCTURE to estimate ancestry for each individual in each dataset. Because the dental data sets had fewer markers and lower total Fisher Information than the genetic datasets, I repeated the analyses on genetics datasets that were subbed to match the number of markers and total Fisher Information of the dental data. Finally, I systematically compared the STRUCTURE results for the genetics and dental datasets. This is the first study to formally assess the suitability of dental data for ancestry estimation, and it is the first study to evaluate the use of commonly used methods from statistical genetics for estimation of ancestry from dental data. It is also the first study to formally compare the utility of dichotomized and non-dichotomized dental data in ancestry estimation.

In Chapter 4, I used ordinally-graded dental morphological data from historic cemeteries and contemporary orthodontics clinics to evaluate variation in African American ancestry across time and geography. I used STRUCTURE and PCA to estimate ancestry, and summarize the trait variation within among the samples. Then I compared the results from the two analyses to make inferences about the causes of trait variation among the African American samples. This study is novel in that it attributes morphological differences among African American populations not only to differences in the amount of African and European ancestry but it also shows that drift either prior to (in African populations) or following (in African American populations in different US regions) admixture played an important role in the differentiation of African American
groups. This finding extends our understanding of the utility of dental morphology for capturing important details about the historical processes that have molded diversity in African Americans.

Chapter 5 is a summary of the findings of the three data analysis chapters and general conclusions.
CHAPTER 2. TESTS OF FIT OF HISTORICALLY-INFORMED MODELS OF AFRICAN AMERICAN ADMIXTURE

This is the peer reviewed version of the following article: Gross JM. 2018. Tests of fit of historically-informed models of African American Admixture. Am J Phys Anthropol 165:211–222, which has been published in final form at https://doi.org/10.1002/ajpa.23343. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Abstract

Objectives:

The African American population in the U.S. formed primarily by mating between Africans and Europeans over the last 500 years. To date, studies of admixture have focused on either a one-time admixture event or continuous input into the African American population from Europeans only. My goal is to gain a better understanding of the admixture process by examining models that take into account (a) assortative mating by ancestry in the African American population, (b) continuous input from both Europeans and Africans, and (c) historically-informed variation in the rate of African migration over time.

Materials and methods:

I used a model-based clustering method to generate distributions of African ancestry in three samples comprised of 147 African Americans from two published sources. I used a log-likelihood method to examine the fit of four
models to these distributions and used a log-likelihood ratio test to compare the relative fit of each model.

Results:

The mean ancestry estimates for my datasets of 77% African/23% European to 83% African/17% European ancestry are consistent with previous studies. I find admixture models that incorporate continuous gene flow from Europeans fit significantly better than one-time event models, and that a model involving continuous gene flow from Africans and Europeans fits better than one with continuous gene flow from Europeans only for two samples. Importantly, models that involve continuous input from Africans necessitate a higher level of gene flow from Europeans than previously reported.

Discussion:

I demonstrate that models that take into account information about the rate of African migration over the past 500 years fit observed patterns of African ancestry better than alternative models. My approach will enrich our understanding of the admixture process in extant and past populations.
Introduction

Anthropologists and geneticists have studied the admixture process in African Americans for decades. These studies typically estimate the proportion of African and European ancestry in African American individuals and populations. Some of these studies attempt to fit simple admixture models to extant patterns of population genetic diversity in African Americans to determine when admixture first began and the per-generation contribution of Europeans (Glass and Li, 1953; Long, 1991; Gravel, 2012; Kidd et al., 2012; Bryc et al., 2015; Baharian et al., 2016). The results of these studies have the potential to assist in uncovering the causes of multifactorial disease and to identify and eliminate the social causes of racial disparity in health outcomes. They also have the potential to help us understand how populations have interacted with one another throughout human history, particularly in cases that involve substantial power asymmetries between the populations.

In terms of ancestry estimates at the population and individual levels, the chief findings of these studies are that (a) the mean African ancestry level in extant African American populations is always substantially higher than the mean European level (Glass and Li, 1953; Long, 1991; Parra et al., 2001a; Bryc et al., 2015), (b) the Native American ancestry level is relatively low throughout the United States (Parra et al., 1998; Tishkoff et al., 2009; Bryc et al., 2015), (c) despite high mean African ancestry proportions at the population level, African ancestry varies substantially among individuals, from a low of about 2% to a high
of about 98% (Bryc et al., 2015), and (d) more mating has occurred between European males and African or African American females than between African American males and European females (Parra et al., 1998, 2001; Bryc et al., 2015).

In terms of the dynamics of the admixture process, studies have produced ambiguous results. Bryc et al. (2015), for example, fit a discrete two-stage admixture model to distributions of ancestry tract lengths and concluded that a single admixture event occurred between Africans and a combine European/Native American group six generations before the present. Baharian et al. (2016) found that a two-pulse admixture model, in which Europeans contributed to the African American population on two separate occasions, fit better than a one-time admixture event model, and that the first incident of admixture occurred around 1740. Glass and Li (1953) assumed a model of continuous gene flow into the African American population from Europeans, and estimated a per-generation rate of European contribution to the African American population of 3.58%. More recently, Parra et al. (2001) concluded that a model of continuous one-way gene flow from Europeans with a per-generation gene flow rate of between 2.0% and 3.1% is consistent with the distribution of European alleles in African American individuals. Even more recently, Jin et al. (2011) fit four admixture models, including a continuous two-way gene flow model, to the distribution of ancestry tracts for a sample of African Americans. The authors constrained their admixture onset time to between 10 and 17 generations in the
past in an effort to be consistent with African American history, and found that one-way gene flow from Europeans into the African American population with an onset of 14 generations in the past produced the best-fit distribution of ancestry tracts.

The results of these studies are broadly incompatible with one another, and, for the most part, they are inconsistent with our current understanding of African American history. Although there is a paucity of historical data about the timing and amount of admixture over the past 500 years, the historical record provides information that can be used to constrain the parameters of admixture models. We know, for example, that the ancestors of African Americans came from diverse locations in Africa, including areas where admixture with people from other regions occurred, such as regions bordering the Mediterranean. However, the vast majority of immigrants both during the slave trade and in recent years came from West Africa (Curtin, 1969; Voyages Database, 2009; US Census Bureau, 2010). We also know that the first importation event to North America involved 20 slaves in Virginia in 1619 (Curtin, 1969). After this initial event, slave importation remained low until the beginning of the 18th century (see Figure 2.1), after which it continued unabated until 1860, even after importation became illegal in 1808 (Curtin, 1969; Smith, 1973). Reports made by slaves discuss forced mating between slaves and slave owners throughout African American history (Federal Writers Project, 2001). Additionally, the passage of anti-miscegenation laws as early as 1664 suggest that African American–
European mating occurred early in U.S. history (General Assembly of Maryland, 1664). More recently, the 2010 Census reports 314,400 immigrants from Northern African and 2,847,199 new immigrants from Sub-Saharan Africa (US Census Bureau, 2010). Based on this information, it is likely that admixture between Europeans and Africans in North America began in earnest as early as 12 generations before the present and that it occurred continuously afterwards. It is also likely that the rate of gene European gene flow into the African American population varied dramatically over the past 400 years as a result of important historical events such as the U.S. Civil War, the passage of anti-miscegenation laws (pre- and post-Civil War), the Great Migration, and the passage of Civil Rights legislation.
Figure 2.1. African migration. The blue points correspond to the vertical scale on the left, the number of African migrants. The solid line corresponds to the vertical axis on the right, the proportion of the African American population these migrants represent (the axis values are rounded to the nearest tenth). This line accounts for both the intrinsic rate of increase, as well as actual numbers of individuals. The dotted line is at 1808, and marks the end of legal slave importation; the points to the left of the dotted line represent people who were forcibly brought to North America to be slaves.

Additionally, records of slave importation indicate that the contributions to African American populations from newly migrated Africans must have been large and persistent (Figure 2.1).Remarkably, to date, with the exception of
those fit by Jin et al. (2011), models of the admixture process have ignored this
African contribution. Jin et al. (2011), however, did not ground the per-generation
source-group contributions in history. Rather, they used observed ancestry
fractions, and divided the contribution from each source population evenly across
the generations in their model.

This study builds on the work of Jin et al. (2012). My novel contributions
include the formal fitting of ancestry models to observed distributions of individual
ancestry in three African American samples, the use of census and other
historical records to vary the model-based per generation contribution of Africans
to the African American population, and the incorporation of assortative mating
by ancestry in the African American population, which I will call “ancestry-related
assortative mating” (AAM). AAM refers to a correlation in ancestry between
mates. Such mating may have been common among African Americans, due, for
example, to geographic structure in the distribution of newly imported slaves, or
selective mating by phenotype among slave owners (Federal Writers Project,
2001). Although there is no historical information about AAM in African
Americans, it has been reported for Hispanic populations in Mexico, the Bay Area
of San Francisco, and Puerto Rico. In Mexican populations in Mexico City and
San Francisco, Risch et al. (2009) found that the correlation in Native American
ancestry between mates was 0.586 and 0.392, respectively. The same authors
found that the correlation in African ancestry between mates was 0.328 in Puerto
Rico. Interestingly, they were unable to identify the social mechanism for AAM.
AAM is relevant to the study of African American admixture because it has the potential to affect estimates of the European contribution to the African American population, a parameter for which we have no direct historical information. By adding AAM to my analyses, I hope to refine our understanding of both the social causes of AAM as well as the European contribution to the African American population.

The goals of this study are to estimate African and European genetic ancestry in three African American samples and to compare the fit of discrete and one-way continuous models of admixture to historically-informed admixture models that incorporate (a) continuous contributions from Africans and Europeans, (b) variable rates of per-generation contribution from Africans, and (c) assortative mating by ancestry in the African American population.

**Methods**

**Data**

I use two datasets for my analyses. The first consists of 1,022,144 autosomal SNP genotypes from 112 Yoruban (YRI), 110 CEPH European (CEU), and 83 adults who self-identified primarily as African American from HapMap Phase 3 (ASW, African ancestry in SW USA) (International HapMap Consortium, 2003). All ASW stated that they had four African American grandparents from the U.S. Southwest.
I filtered the ASW sample to remove related individuals. For this step, I excluded children from parent/offspring trios and duos, resulting in a sample of 49 unrelated individuals. I filtered the SNPs in two ways. First, I retained SNPs that were common to the YRI, CEU, and ASW. This step reduced the number of SNPs to 992,601. Second, following the work of Pfaff et al. (2004), I limited my analyses to SNPs that were informative about the admixture process. Those authors showed that marker informativeness, captured by Fisher Information (FI), is a function not only of differences in allele frequencies between putative parental populations, but also of the allele frequencies themselves. Specifically, I used an FI cutoff of 2.5, which, while arbitrary, eliminated thousands of uninformative loci without inflating the error in individual ancestry estimates (see below). Data filtration was conducted in R.

The second sample consists of 645 autosomal short tandem repeat genotypes from 50 Yoruba, 29 French, and 98 self-identified African American individuals from four locations in the U.S. Midwest and East Coast: Baltimore, Chicago, North Carolina and Pittsburgh (Tishkoff et al., 2009); I refer to this sample as African Americans in the Midwest and East Coast (AME) I performed my analyses on both the full AME sample and separately on the Baltimore sample of 44 individuals (ABT). The Baltimore sample is the largest of the African American samples in the Tishkoff et al. (2009) dataset. I analyzed this sample separately to control for the possibility that the AME sample is structured with respect to ancestry simply because it is comprised of individuals from multiple
geographic locations, each of which may have experienced a different admixture history.

The YRI and CEU samples served as parental source populations for analyses of the ASW, and the Yoruba and French samples served as parental source populations for analyses of the AME and ABT samples. I recognize that these samples are not the true parental source populations, which derived from diverse locations in Europe and Africa (Montinaro et al., 2015; Patin et al., 2017). For this reason, my individual-level ancestry estimates, and the mean population-level estimates, are unlikely to be accurate. This limitation is common to all admixture studies. In my study, this limitation could affect my estimates of the per generation contribution of the Europeans and Africans to the admixed population (see below). However, the absence of true parental sources is unlikely to affect the shape of the observed ancestry distributions because the error would be systematic, and therefore unlikely to change my conclusions about the relative fit of different models of the admixture process.

Table 2.1. Sample sizes and number of loci in the three African American samples

<table>
<thead>
<tr>
<th>Sample</th>
<th>n</th>
<th>Loci</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASW</td>
<td>49</td>
<td>7,392 SNPs</td>
</tr>
<tr>
<td>AME</td>
<td>98</td>
<td>645 STRs</td>
</tr>
<tr>
<td>ABT</td>
<td>44</td>
<td>645 STRs</td>
</tr>
</tbody>
</table>
Ancestry estimation and model construction

My strategy for estimating African and European ancestry and for comparing different admixture models consisted of three steps. First, I estimated individual-level African and European ancestry in the ASW, AME, and ABT samples using the Bayesian model-based clustering algorithm implemented in the program STRUCTURE (Pritchard et al., 2000). For each sample, I ran STRUCTURE five times at \( K = 2 \) through \( K = 6 \) using a burnin phase of 25,000 steps and 15,000 MCMC repetitions. Otherwise, I used the default settings in STRUCTURE. Second, I fit four models to the distributions of individual African ancestry (see Table 2.2).

Table 2.2. Model parameters each generation

<table>
<thead>
<tr>
<th>Model</th>
<th>( g_0 )</th>
<th>( g_{1+} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. One-time admixture event</td>
<td>( \alpha, \beta )</td>
<td>-</td>
</tr>
<tr>
<td>2. One-way gene flow</td>
<td>( \alpha, \beta )</td>
<td>( \alpha, \beta )</td>
</tr>
<tr>
<td>3. Two-way gene flow</td>
<td>( \alpha, \beta )</td>
<td>( \alpha, \beta )</td>
</tr>
<tr>
<td>4. two-way + AAM</td>
<td>( \alpha, \beta, R )</td>
<td>( \alpha, \beta, R )</td>
</tr>
</tbody>
</table>

Figure 2.2 shows a generalized version of the four models. Table 2.3 shows the ancestry of possible mating pairs in each generation of Models 3 and 4 after the first generation.

European x European matings were excluded from the models. Thus, under this mating scheme, all individuals in the African American population at any given time must have at least one African ancestor.
Figure 2.2. Generalized admixture model. A = African (green circles), E = European (yellow circles), AA = African American (blue circles). The African American population forms at $g_1$, from the fractional contributions of the European and African source populations, $\alpha_0$ and $\beta_0$, at time $g_0$. $R$ is the correlation in ancestry between mates, which affects the mating pairs in the red box.

Three parameters were associated with each model: the number of generations, $g$, since the onset of admixture, and the contributions from the European and African source populations $\alpha$ and $\beta$, respectively. A fourth parameter, $R$, the correlation in ancestry between mates, was associated with Model 4.
Table 2.3. Possible mating pairs after generation 1 in each model

<table>
<thead>
<tr>
<th>Mating pairs</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>African American x African American</td>
<td>1 - 4</td>
</tr>
<tr>
<td>African American x African</td>
<td>3 and 4</td>
</tr>
<tr>
<td>African x African</td>
<td>3 and 4</td>
</tr>
<tr>
<td>African American x European</td>
<td>2 - 4</td>
</tr>
<tr>
<td>European x African</td>
<td>3 and 4</td>
</tr>
</tbody>
</table>

For all models, I assumed that admixture began \( g = 12 \) generations ago, in approximately 1700, when there was a surge in slave importation. For the one-time admixture event model (Model 1), the parameters \( \alpha \) and \( \beta \) were set to zero after the initial admixture event at generation \( g_0 \). Under this model, \( \alpha_0 \) and \( \beta_0 \) are equal to the ancestry fractions in the current African American samples (ASW, AME, and ABT). For the one-way gene flow model, \( \alpha \) was set to a constant rate per generation, \( \beta_0 \) was set to \( 1 - \alpha_0 \), and \( \beta \) was set to zero for all subsequent generations. For the two-way gene flow models, I used the African slave import estimates from Curtin (1969) to approximate the per-generation African contribution, \( \beta \), during the slave trade. These estimates are based on English slave trade data and estimates of slave trade importation (Curtin, 1969); thus, the rates change each generation. I used information from the U.S. Census (Gibson and Lennon, 1999) on native-born African Americans vs. African immigrants to determine the \( \beta \)-values for each generation subsequent to the cessation of legal slave importation.
Generating model-based individual African American ancestry distributions

For each model, I produce an expected distribution of individual African ancestry (IA) as follows. My measure of African ancestry is an individual’s number of African ancestors in a given generation. For a given model, the IA distribution for the first generation was formed by “mating” Europeans and Africans in the proportions $\alpha$ and $\beta (1 - \alpha)$. In the first generation of existence, $g_1$, individuals in this newly formed African American population could have either 50% or 100% African ancestry, that is, either one or two African ancestors. In subsequent generations, IA distributions were created by drawing mating pairs from the IA distribution in the previous generation in accord with the relevant model parameters (Tables 2 and 3). Thus, for all models, at any given generation, African American individuals had any of a discrete number of African ancestors, corresponding to African ancestry proportions between $1/(2^g)$% and 100%.

For Models 1–4 for each sample, the parameter $a$ was set to values that produce the observed mean European ancestry in that sample after 12 generations. To achieve this outcome for Model 1, $\alpha$ was zero after generation 1. For Models 2–4, $\alpha$ was constant across generations.
Ancestry-related assortative mating

For Model 4, I used a Monte Carlo simulation approach to incorporate AAM into my mechanistic admixture model. For each simulation, I began by creating a vector of 20,000 pairs of individual ancestry estimates, with each estimate being drawn at random (with replacement) from the previous generation’s IA distribution. In this way, I effectively created a series of randomly mating couples. I then (a) calculated the correlation in ancestry between mates in the vector, (b) permuted one member of each of two mating pairs, (c) recalculated the correlation for the newly permuted vector, and (d) compared the new correlation to original correlation. At each step in the process, I retained permuted vectors that had a correlation in ancestry between mates that was closer to my target correlation, and reverted back to the previous vector when the permutation produced a correlation that was further from my target. In this way, I am able to create new versions of Models 1–3 that incorporated correlations in ancestry between mates ($R$) ranging from 0 to 1 in increments of 0.01.

African contribution to the African American population

In Models 3 and 4, which include continuous African gene flow into the African American population, for each generation during legal slave importation, the proportion of immigrant Africans was calculated as the number of imported slaves divided by the sum of the number of imported slaves including and prior to that generation. The post-1850 data were taken from the U.S. Census Bureau.
(Gibson and Lennon, 1999), which reports the number of native-born African Americans and the number of African immigrants in 10-year increments beginning in 1850. Since I used a generation time of 25 years, I calculate population sizes for each generation by combining population data for two 10-year census reports, then adding the midpoint value from a third report. $\beta$-values were then calculated as the number of African immigrants for each generation divided by the total African American population size according to the U.S. Census.

**Changes in the IA distributions over time**

I iterated each of the four models until the expected distributions reached a steady-state to better understand the change in the expected IA distribution over time. For the one-time admixture event model (Model 1), I began with contributions $\alpha = 78\%$ European and $\beta = 1 - \alpha = 22\%$ African. For the one-way gene flow model (Model 2), I set $\alpha$ at 0.05 per generation. For the models involving two-way gene flow (Models 3 and 4), $\alpha$ was set at a constant-rate of 0.05, and $\beta$ was set according to slave importation and U.S. Census data for the first 12 generations, then set to $\alpha$ constant rate of 0.05 for all remaining generations.
Testing the fit of the models to the observed distribution of African American ancestry

I tested the fit of each model-based distribution to the IA distribution for each of the three samples (ASW, AME, and ABT). Based on historical information described in the introduction, I assumed that admixture began 12 generations ago, and I compared the fit of each model at this 12-generation point. My method for comparing the fit of the models was as follows.

Because individuals in admixed populations have a discrete number of ancestors from the parental source populations, for my statistical tests of fit, I first divided the individuals in each observed sample into 16 bins from 1/(2^9)\% to 100\% African ancestry. Similarly, for each model, I binned the model-based probabilities into 16 bins from 1/(2^9)\% to 100\% African ancestry.

I calculated the log-likelihood of multinomial cell probabilities for the IA distribution produced by each model using Equation 2.1,

\[
\ln (L) = \sum_{i=1}^{m} x_i \ln (p_i) \tag{Eq. 2.1}
\]

where \(m\) is the number of bins in the distribution (\(m = 16\)), \(x_i\) is the number of individuals in bin \(i\) in the observed distribution, and \(p_i\) is the probability of the number of African ancestors in bin \(i\) in the expected distribution.

After calculating the log-likelihood for each of the expected distributions, I assessed their fit relative to one another using a likelihood ratio statistic, LLR (Equation 2.2) (Sokal and Rohlf, 2012).

\[
\text{LLR} = -2(\ln(L_0) - \ln(L_1)) \tag{Eq. 2.2}
\]
For large samples, the distribution of LLR is approximated by the $\chi^2$ distribution. In each case, I used the best-fit distribution of the previous model as the null hypothesis ($\ln(L_0)$).

Results

Ancestry estimates

All but 7,392 of the SNPs in the full HapMap dataset had FI values below 2.5 (see Figure A.1). This result is unsurprising given the relatively young age of our species and the fact that a small amount of the total variation in our species is unique to populations and regions (Rosenberg et al., 2002). Nonetheless, the remaining 7,392 loci were highly informative, resulting in average individual-level error estimates of $\pm 0.6\%$ (based on the 95% credible regions estimated in STRUCTURE).

My STRUCTURE runs showed average individual-level Native American ancestry estimates of 0.08% at $K = 3$, 0.4% at $K = 5$, and 0.6% at $K = 6$ for the Tishkoff et al. (2009) dataset, which includes the AME and ABT samples. The HapMap dataset does not include a Native American sample. However, previous studies have shown the average Native American ancestry to be <1% among African Americans across the United States, and <2% in African Americans in the U.S. Southwest (Bryc et al., 2015, Jin et al., 2012). Based on these results, I conduct my analyses under the assumption of dihybrid ancestry.
The mean and range of African American ancestry estimates for each sample are listed in Table 2.4. The mean estimates are consistent with those from previous studies of African American ancestry (Glass and Li, 1953; Parra et al., 1998; Falush et al., 2003; Oksenberg et al., 2004; Tishkoff et al., 2009; Bryc et al., 2010, 2015; Kidd et al., 2012).

Table 2.4. Mean and range of African American ancestry estimates for the ASW, AME, and ABT samples

<table>
<thead>
<tr>
<th>Sample</th>
<th>Mean African ancestry</th>
<th>Individual African ancestry range</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASW</td>
<td>76.53%</td>
<td>58.43% - 91.74%</td>
</tr>
<tr>
<td>AME</td>
<td>80.59%</td>
<td>45.25% - 98.40%</td>
</tr>
<tr>
<td>ABT</td>
<td>83.06%</td>
<td>62.60% - 98.40%</td>
</tr>
</tbody>
</table>

The observed IA distribution for each sample are shown in Figure 2.3. Two predominant features of all three distributions are a relatively high mean African ancestry level and a left skew. Differences between the distributions include an absence of individuals in the highest ancestry bin in the ASW sample, and a greater left skew towards low African ancestry in the AME sample compared with the other two samples.
Figure 2.3. Individual ancestry distributions. (A) ASW, (B) AME, (C) ABT

Model expectations over time

Figure 2.4 shows the change in the expected IA distribution for each model at 4, 12, 75, and 125 generations after the onset of admixture. I chose to begin at four generations because it produced the highest likelihood under Model 1, the one-time admixture event model, when fit to all three samples. Under this model, variation in African ancestry is lost each generation; by 20 generations, everyone in the population has a single African ancestry value that is equivalent to the original contributions from Africans and Europeans. Under Model 2, continuous one-way gene flow from Europeans and no contribution from newly migrated Africans, variation in African ancestry is reduced each generation until, by 125 generations, everyone has 0% African ancestry. Under Models 3 and 4,
involving continuous gene flow from both Europeans and Africans, variation in 
African American ancestry is maintained among individuals over time, and the 
distribution eventually comes to a steady-state by 75 generations.

It is important to note that, even though the IA distributions for the models 
differ from one another over time, in some cases markedly, with the exception of 
Model 1, the expected IA distributions are similar at 12 generations. These 
results suggest that my statistical tests of fit will lead to a clear rejection of Model 
1, but it may be considerably more difficult to distinguish between the remaining 
models.
Figure 2.4. Expected model distributions over time. (a) One-time admixture event, (b) One-way gene flow, (c) Two-way gene flow, (d) Two-way gene flow with AAM. The choice of values for the generations were based on the best fit (highest likelihood) for Model 1 (4 generations), historical information about the onset of admixture (12 generations), the time to steady state at 80% African ancestry for Model 1 (20 generations), the time to steady state for Models 3 and 4 (75 generations), and the time to steady state at 0% African ancestry for Model 2 (125 generations).
Model comparisons

Table 2.5 shows the log-likelihoods for each model (row 1), LLRs comparing the fit of the models (row 2), $p$-values for the LLR tests (row 3), $\alpha$-values for Models 2–4 (row 4), and $R$-values for Model 4 (row 5). For all three samples, as predicted from the IA distributions in Figure 2.4, Model 1 has by far the lowest log-likelihood (poorest fit), and Model 2 fit better than Model 1 (significantly higher log-likelihood at $p < 0.05$). For Models 2–4, the $\alpha$-values (European contribution) that produced the highest log-likelihoods ranged from 0.037 (ABT, Model 2) to 0.071 (ASW, Model 4). The $R$-values for Model 4 ranged from 0.01 for the ASW sample to 0.15 for the ABT sample.
Table 2.5. Model-fitting results for the ASW, AME, and ABT samples at \( g = 12 \) generations

<table>
<thead>
<tr>
<th></th>
<th>ASW</th>
<th>Model 1: One-time Admixture</th>
<th>Model 2: One-way gene flow</th>
<th>Model 3: Two-way gene flow</th>
<th>Model 4: Two-way + AAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \ln L )</td>
<td>-526.253</td>
<td>-81.425</td>
<td>-84.326</td>
<td>-83.838</td>
<td></td>
</tr>
<tr>
<td>( LLR )</td>
<td>-</td>
<td>889.656*</td>
<td>-5.802*</td>
<td>0.974</td>
<td></td>
</tr>
<tr>
<td>( p )</td>
<td>-</td>
<td>1.74E-195</td>
<td>0.016</td>
<td>0.324</td>
<td></td>
</tr>
<tr>
<td>( \alpha )</td>
<td>-</td>
<td>0.051</td>
<td>0.068</td>
<td>0.071</td>
<td></td>
</tr>
<tr>
<td>( R )</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AME</td>
<td>One-time Admixture</td>
<td>One-way gene flow</td>
<td>Two-way gene flow</td>
<td>Two-way + AAM</td>
</tr>
<tr>
<td>( \ln L )</td>
<td>-1369.914</td>
<td>-335.781</td>
<td>-190.56</td>
<td>-188.435</td>
<td></td>
</tr>
<tr>
<td>( LLR )</td>
<td>-</td>
<td>2068.266*</td>
<td>290.444*</td>
<td>4.251*</td>
<td></td>
</tr>
<tr>
<td>( p )</td>
<td>-</td>
<td>0</td>
<td>3.98E-65</td>
<td>0.039</td>
<td></td>
</tr>
<tr>
<td>( \alpha )</td>
<td>-</td>
<td>0.042</td>
<td>0.057</td>
<td>0.059</td>
<td></td>
</tr>
<tr>
<td>( R )</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ABT</td>
<td>One-time Admixture</td>
<td>One-way gene flow</td>
<td>Two-way gene flow</td>
<td>Two-way + AAM</td>
</tr>
<tr>
<td>( \ln L )</td>
<td>-458.48</td>
<td>-184.888</td>
<td>-83.263</td>
<td>-80.832</td>
<td></td>
</tr>
<tr>
<td>( LLR )</td>
<td>-</td>
<td>547.184*</td>
<td>203.251*</td>
<td>4.861*</td>
<td></td>
</tr>
<tr>
<td>( p )</td>
<td>-</td>
<td>5.16E-121</td>
<td>4.08E-46</td>
<td>0.027</td>
<td></td>
</tr>
<tr>
<td>( \alpha )</td>
<td>-</td>
<td>0.037</td>
<td>0.05</td>
<td>0.057</td>
<td></td>
</tr>
<tr>
<td>( R )</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.15</td>
<td></td>
</tr>
</tbody>
</table>

\( \ln L \), log likelihood; \( LLR \), log likelihood ratio; \( p \), \( p \)-value; \( \alpha \), best-fit European contribution; \( R \), best-fit correlation in ancestry between mates.

*\( p < .05 \).

**Sample-specific results**

Figure 2.5 shows the expected distributions for the best-fitting versions of Models 1–4 (at 12 generations) in pink and the observed IA distribution in pale blue for ASW. Figure 2.5a shows that Model 1 fits poorly because the observed ASW IA distribution retains substantial variation in African ancestry compared with that predicted under a one-time model. Model 2 (Figure 2.5b), involving continuous gene flow from Europeans, maintains substantially more variation in African ancestry after 12 generations, and therefore fits substantially better than Model 1. The differences in fit between the remaining models are more subtle. Model 3, involving continuous gene flow from both Europeans and Africans fits...
significantly worse than Model 2. Adding AAM (Model 4, Figure 2.5d), however, did not result in a difference in fit relative to Model 3. One noteworthy feature of the expected IA distributions for Models 2–4 is the slight swell in frequency in the left tail of the distribution around 0.44. This swell is absent from the observed ASW distribution. I return to this issue in the discussion.

Figure 2.5. ASW IA distributions for 12 generation models. (A) One-time admixture, (B) One-way gene flow, (C) Two-way gene flow, (D) Two-way gene flow with AAM. Model distributions are shown in pink, and the observed distribution is shown as transparent blue bars with thick black outlines.

For the AME sample (Figure 2.6), Model 1 again fits poorly, and Model 2 fits significantly better than Model 1. In contrast to the ASW sample, Model 3 fits significantly better than Model 2, and Model 4, with AAM, fits significantly better
than Model 3. The correlation in ancestry between mates, $R$, that produced the best fit for Model 4 was 0.08. Again, the expected distributions for Models 2–4 have a slight swell in frequency around 0.44 that is absent in the observed IA distribution.

Figure 2.6. AME IA distributions for 12 generation models. (A) One-time admixture, (B) One-way gene flow, (C) Two-way gene flow, (D) Two-way gene flow with AAM. Model distributions are shown in pink, and the observed distribution is shown as transparent blue bars with thick black outlines.

The pattern of fit for the ABT sample, shown in Figure 2.7, is the same as that for the AME. In this case, the correlation in ancestry that produced the best
fit for Model 4 was 0.15, almost twice as high as it was for the AME. Again, the observed IA distribution lacks the slight swell in frequency around 0.44.

Figure 2.7. ABT IA distributions for 12 generation models. (A) One-time admixture, (B) One-way gene flow, (C) Two-way gene flow, (D) Two-way gene flow with AAM. The model distributions are shown in pink, and the observed distribution is shown as transparent blue bars with thick black outlines.

Discussion

In this study, I fit four admixture models to the distribution of IA in three samples of American Americans. The models took into account historical information about immigration. In all models, the admixture process began 12 generations ago, approximating the onset of high rates of slave importation into
North American beginning around 1700 (Curtin, 1969). Models 2 - 4, in accord with slave reports, other historical accounts, and U.S. Census data (Gordon-Reed, 1998; Federal Writers Project, 2001), incorporated ongoing gene flow from Europeans. Models 3 and 4 incorporated ongoing gene flow from newly immigrated Africans in accord with historical data on slave importation (Curtin, 1969) and the U.S. Census Bureau (Gibson and Lennon, 1999). To my knowledge, to date, no study of genetic admixture has incorporated ongoing African input into the African American population, despite the fact that it may have been the most prevalent source of gene flow into the African American population for generations after the onset of slavery (see Figure 2.1) (Curtin, 1969; Gibson and Lennon, 1999). To my knowledge, this is also the first study to incorporate AAM into models of the admixture process in a population. Although there is no historical information about assortative mating by ancestry in African Americans, it has been documented in Hispanic populations (Risch et al., 2009).

I found that, in all cases, the one-time admixture event model fit the observed IA distributions poorly relative to other models, which involve ongoing gene flow from one or both source populations. The better fit is, in part, a result of the fact that ongoing gene flow maintains variation in ancestry in admixed populations; this variation erodes rapidly under a model involving a single admixture event (Verdu and Rosenberg, 2011). A one-time admixture event model is also inconsistent with slave narratives that describe rape by white slave owners and abortion attempts for the potential offspring of European-
African/African American unions (Federal Writers Project, 2001), and it inconsistent with data from the U.S. Census Bureau. “Mulatto”, for example, was a category on the U.S. Census from 1850 to 1890 and again from 1910 to 1920. More recently, the U.S. Census has allowed individuals to choose multiple races. According to the 2010 U.S. Census, 9.0 million U.S. residents reported multiple-race ancestry (Jones and Bullock, 2012). Of these, about 2.3 million individuals chose combinations involving white and black race. These data may underreport mixed race ancestry in the United States; an independent analysis of census data by the Pew Research Center (Parker, et al., 2015) found that 6.9% of American residents had multiple race origins, as opposed to the 2.1% identified on 2013 American Community Survey. Furthermore, in 2014, according to the U.S. Census Bureau, 7% of African American men were married to European American women, and 4% of African American women were married to European American men (Current Population Survey, 2014). This marriage rate is consistent with the per-generation European gene flow rate (a) in my best-fit one-, two-, and two-way gene flow with AAM models.

Although these data and my results are inconsistent with a one-time admixture event model, they are at odds with findings from recent high-profile genetic studies (Gravel, 2012; Bryc et al., 2015; Baharian et al., 2016). Recently, Bryc et al. (2015), for example, fit a two-event model of admixture to a large, nation-wide sample of African American genetic data. The best fit-version of this model involved a one-time admixture event between Native Americans and
Europeans 12 generations ago followed by a one-time admixture event between this group and Africans six generations ago. This result makes sense in terms of the amount of variation in ancestry in African Americans. One-time admixture events that occurred earlier would result in less variation, and one-time events that occurred later (more recently) would result in a wider range of variation (Verdu and Rosenberg, 2011). However, this result does not make sense in terms of the shape of the IA distribution. One time models produce symmetric IA distributions (as do continuous models with equal contributions from both parental sources). Neither the distribution from Bryc et al. (2015) nor the observed distributions for the ASW, AME, and ABT samples are symmetric; they all have strong skews toward low African ancestry. My modeling results (Figure 2.4), as well as those from Verdu and Rosenberg (2011), demonstrate that skewed distributions are the result of asymmetric contributions from parental sources under continuous gene flow models (including zero contribution from one of the parental sources). Furthermore, this one-time admixture event model, and, for that matter, any one-time admixture event model, is inconsistent with census and other historical records documenting mating between Americans of African and European descent. Based on these results, I reject the hypothesis of one-time admixture in African Americans.

The history of continuous slave importation from about 1700 to 1860, as well as the continued post-slavery migration of Africans to what is now the United States, led us to predict that the two-way gene flow model would fit the three
observed IA distributions better than one-way gene flow model. I am therefore surprised by the finding for the ASW sample that the two-way gene flow model fit significantly worse than the one-way model. It is possible that African Americans in the U.S. Southwest were relatively isolated from African Americans along the east coast beginning in the 19th century; this result may be consistent with such a history. However, the lack of fit could also reflect the sampling scheme used to collect the ASW data. This scheme excluded individuals with any non-African American parent or grandparent (International HapMap Consortium, 2003), that is, it excluded individuals with African and/or European ancestors in the previous two generations. This sampling scheme could explain the low average African ancestry in the ASW sample relative to the AME and ABT (Table 2.4), and it could explain the fact that the ASW sample was the only one of the three to lack individuals with >98% African ancestry. These results highlight the importance in studies of admixture models of collecting representative samples of admixed populations.

In contrast, as expected from the historical and census data, Model 4, with two-way gene flow and AAM, fit best for the AME and ABT samples. An important finding of this study is that input from Africans into the African American population necessitates a concomitant increase in the per-generation contribution from Europeans compared with a one-way gene flow model. AAM also necessitates a higher contribution from Europeans compared with models lacking AAM. For example, for the ABT sample, the per-generation European
contributions under the one-, two-, and two-way with AAM models respectively were 3.0%, 5.0%, and 5.7%. The values were even higher for the ASW and AME samples. These results imply that the per-generation contributions to African American populations from Europeans may have been in excess of 5% throughout U.S. history.

No information is provided about the sampling scheme for the AME sample (of which the ABT is a subset). However, the pattern of the lack of fit, an observed excess of African ancestry in the highest bins and a deficit in the lowest, may indicate an absence of recent European contribution, which is inconsistent with census data but consistent with a sampling scheme that excluded individuals with recent European ancestors.

Along the same line, none of the observed distributions had a slight swell in frequency around 0.44. This slight swell was seen in all of the models that included continuous gene flow from Europeans. These swells are actually distributions that are produced by matings between African Americans and Europeans each generation. The same phenomenon occurs with continuous contribution from Africans to the African American population; however, these distributions are not as apparent because they are contained within main distribution. The absence of this feature in the observed IA distributions reflects a lack of European contribution in recent generations because there was no European contribution, because individuals with recent European ancestry do not
self-identify as African American, or because the sampling scheme excluded recent European ancestors.

Independent of the sampling scheme, there are other possible reasons for residual lack of fit of the models to the observed distributions. These reasons include the failure of my relatively simple models to capture the true complexity of African American history. My models, for example, do not take into account the potential effects of population substructure due to processes other than AAM, for example, the AME sample is comprised of individuals from four locations that may have experienced limited gene flow, or whose ancestors came from different places. Such substructure has led to heterogeneity in the distribution of African ancestry among African Americans in different regions of the country (Bryc et al., 2015). Other possible reasons for the lack of fit include reduced power associated with low sample sizes and limitations of the ancestry-estimation methods, for example, a lack of correct source populations.

Additionally, I did not include Native American contributions in my models. This choice was justified in part by the fact that the Native American contribution is low. My STRUCTURE analyses showed the mean Native American ancestry to be below 1% for values of $K$ between three and six, and according to the large-scale analyses of thousands of African Americans by Bryc et al. (2015), the mean Native American ancestry among African Americans is 0.8%. Although the average ancestry proportion was higher in the west and southwest, it was still $<2%$. Furthermore, Jin et al. (2012) found only negligible amounts of ancestry
from groups other than Europeans and Africans in their sample of 1,890 African Americans. Importantly, the failure to include Native American ancestry would not affect the fit of one- and two-way gene flow models unless contributing African and European sources themselves contained substantial Native American ancestry prior to mating with individuals in the African American population. Even in this case, for which there is no historical evidence, the shapes of the model-based distributions and the pattern of lack of fit would not be affected.

Conclusions

I conclude that (a) admixture models that are informed by our understanding of African American history fit better than simplistic models involving one-time admixture events, (b) historically-informed models suggest that the European contribution to African American populations has been higher than previously reported, (c) future studies of the admixture process should collect representative samples of admixed populations, and (d) future studies of the admixture process may benefit from exploring AAM.
CHAPTER 3. INFORMATIVENESS OF DENTAL MORPHOLOGY IN ANCESTRY ESTIMATION IN AFRICAN AMERICANS

This is the peer reviewed version of the following article: Gross JM, Edgar HJH. 2019. Informativeness of dental morphology in ancestry estimation in African Americans. Am J Phys Anthropol 168:521–529, which has been published in final form at https://doi.org/10.1002/ajpa.23768. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Abstract

Objectives:

Our objective is to assess the informativeness of dental morphology in estimating biogeographic ancestry in African Americans.

Materials and methods:

The data are 62 dental morphological traits scored as nondichotomized and dichotomized in 797 individuals, 992,601 SNPs from 271 individuals, and 645 STRs from 177 individuals. Each dataset consists of Africans, Europeans, and African Americans. For each dataset, we summed Fisher Information (FI), then used STRUCTURE to estimate ancestry.

Results:

Total FI was highest for SNPs, followed by STRs, nondichotomized dental traits, and dichotomized dental traits. For both genetic datasets, Africans and
Europeans fell into two distinctive clusters with low 90% credible regions for individual ancestry estimates. In African Americans, membership in the African cluster was 76.4% and 80.4% for SNPs and STRs, respectively. For the dental data, all Africans and Europeans had appreciable membership in both clusters and comparatively high 90% credible regions for individual ancestry estimates. Nonetheless, African Americans had consistently higher membership in the same cluster in which Africans had high membership. African American membership in this cluster was significantly higher for the nondichotomized form than for the dichotomized.

Discussion and conclusions:

FI potentially provides a useful gauge of the effectiveness of dental and genetic data for ancestry estimation. The comparatively high FI of nondichotomized dental traits suggests data in this form may be better suited for studies of admixture than dichotomized data. Because of high error in individual ancestry estimates, dental morphological data may be unable to distinguish differences in ancestry among individuals within admixed populations.

Introduction

Both metric and nonmetric cranial morphological traits have long been used to study human prehistory (Berry and Berry, 1967; Ossenberg, 1976; Jose et al., 2001; Hanihara et al., 2012; Movsesian, 2013). Morphology has been
shown to be heritable and therefore reflective of genetic diversity (Sjøvold, 1984; Velemínský and Dobšíková, 2005). Numerous studies have shown global concordance between patterns of cranial morphology and neutral genetic diversity. Roseman (2004), for example, found that differences in cranial morphology among populations mirror differences in neutral genetic markers. Other studies have found a positive correlation between cranial morphological diversity and geographic distance that recapitulates patterns found with neutral genetic markers (Hanihara et al., 2003; Relethford, 2004). More recently, Betti, Balloux, Amos, Hanihara, and Manica (2009) showed that the level of cranial morphological diversity within populations decreased steadily with increasing geographic distance from Africa, reflecting a serial founder effect process as humans spread from Africa during the Pleistocene (Ramachandran et al., 2005). Overall, the results of these studies indicate that cranial morphology is a potential proxy for neutral genetic diversity and therefore appropriate for use in studies of evolution.

Similar work has been done using dental morphology. Following the establishment of the Arizona State University Dental Anthropology System (ASUDAS) in the 1970s, research that characterized and compared dental morphology among populations flourished (Turner, 1984; Irish, 1997, 1998a; b; Scott and Turner, 1997). Consistent with the results obtained from cranial morphology and neutral genetic data, this research has identified positive correlations between global morphological diversity and geographic distance
(Irish and Guatelli-Steinberg, 2003), as well as patterns of within and between regional variation that are consistent with a serial founder effect process (Hanihara, 2008). Researchers have also used multivariable analyses of dental morphology to assign individuals to locations of origin in different world regions (Edgar, 2005, 2015; Irish, 2015). Overall, these studies indicate that dental morphology has the potential to provide important insights into the evolutionary processes that have molded human diversity.

Recently, anthropologists have begun to use methods from statistical genetics to examine admixture between previously separated populations using morphological data. Algee-Hewitt (2016) used a cluster-based method to estimate ancestry from cranial morphology and compared her results to those from published genetic studies. The goal of this study was to test whether biogeographic ancestry could be reliably estimated using cranial morphology. She found that the results of her analyses corresponded to those in the published studies in terms of patterns of population structure and patterns of admixture.

Our primary goal in this study is to determine whether a similar approach can be applied to dental morphology. Specifically, we wish to determine whether a model-based clustering method that is frequently applied to genetic data can be used to distinguish African, European and African American populations from one another, and whether the method can be used to distinguish different levels of African and European ancestry in admixed African American individuals. A secondary goal is to determine whether ordinally-graded dental morphological
traits perform better in ancestry estimation than dichotomized versions of the same traits.

To accomplish these goals, we scored dental morphological traits in historic and contemporary Africans and Europeans, and contemporary African Americans. The traits were first scored as ordinally-graded character states and then dichotomized into categories using breakpoints (Edgar, 2017). To assess the potential of the data to be informative about admixture, we computed Fisher Information for each dental trait, following Pfaff, Barnholtz-Sloan, Wagner, and Long (2004). We then used the model-based clustering method implemented in the program STRUCTURE (Pritchard et al., 2000) to assign fractional ancestry of individuals to each of two genetic clusters. When an admixed population is formed from two relatively isolated “source” populations, in STRUCTURE, individuals in one source population will have approximately 100% membership in one genetic cluster, and individuals in the other source population will have approximately 100% membership in a second genetic cluster. In contrast, individuals in the admixed population will have fractional membership in both clusters. For comparative purposes, we performed the same analyses on two published genetic datasets, one consisting of 992,601 autosomal single nucleotide polymorphisms (SNPs) for 112 Africans, 110 Europeans, and 49 African Americans, and another consisting of 645 autosomal short tandem repeat genotypes (STRs) from 50 Yoruban, 29 French, and 98 self-identified African American individuals from four locations in the United States.
Materials and methods

Data

We make the following assumptions: (a) African American populations were formed in the United States over the last several hundred years through mating between sub-Saharan African and European source populations, (b) admixture is the sole determinant of allele frequencies in the admixed population, that is, drift and selection have not affected allele frequencies; (c) dental traits are reflective of underlying genetic variation.

Many Africans immigrated to the United States from diverse locations across the African continent whether forcibly during slavery or, after the 19th century, by choice. However, slave importation records and US Census Bureau data show that the vast majority of Africans came to the United States from West Africa (Curtin, 1969; Gibson and Lennon, 1999; US Census Bureau, 2010). We therefore use genetic and dental data from West Africans in our analyses.

We use one dental morphological dataset and two genetic datasets for our analyses. The dental data are 62 morphological traits from 196 West Africans, 335 Europeans and European Americans, and 266 African Americans from New York, South Carolina, Tennessee and Washington (Table A.1.). All African American samples came from orthodontic collections in these locations; their race was assigned by the orthodontists who assembled the collections. All dental data were collected by one of the authors (HJHE). The traits were first scored as graded (Turner II et al., 1991) and then dichotomized using breakpoints following
Edgar (2017). The analyses described below were applied separately to both the dichotomized and graded (hereafter referred to as “nondichotomized”) forms of the dataset. We eliminated traits scored in less than 50 individuals per population and individuals with more than 50% missing data from further consideration. This resulted in a set of 53 traits from 168 Africans, 295 Europeans, and 256 African Americans with an average of 18% missing data per trait.

The two genetic datasets were obtained from published sources (Table A.2.). The first genetic dataset consists of 992,601 autosomal SNPs for 112 Africans, 110 Europeans, and 49 individuals from the US Southwest who identified primarily as African American and had four African American grandparents who were also born in the Southwest (International HapMap Consortium, 2003). The second dataset consists of 645 autosomal STRs from 50 Yoruban, 29 French, and 98 self-identified African American individuals from four locations in the US Midwest and East Coast: Baltimore, Chicago, North Carolina, and Pittsburgh (Tishkoff et al., 2009). The SNPs have an average of 0.4% missing data per marker, and the STRs have an average of 3% missing data per marker. These sampling locations differ from the sampling locations for the dental data. This distinction is relevant because African and European ancestry proportions in African Americans vary from region to region (Bryc et al., 2010; Baharian et al., 2016).
To facilitate our analyses, we converted the data to numeric format: we used 1–4 for the four nucleotides for the SNPs, PCR fragment lengths for the STRs, and trait scores for the dental morphological data.

**Fisher Information**

To be informative about ancestry in an admixed population, a genetic marker must differ in frequency between the source populations that contributed to it. The difference in allele frequency between source populations is referred to using the symbol $\delta$. However, Pfaff et al. (2004) showed that the informativeness of a genetic marker also depends on the allele frequencies in the source populations independent of $\delta$, as well as the relative contribution of the sources to the admixed population. Following Pfaff et al., we calculated trait/locus specific and total Fisher Information (FI) for the full set of markers in each dataset as:

$$
E[J(m_1)] = \sum_g \sum_k \frac{\delta_{g,k}^2}{\hat{p}_{g,Ak}}
$$

Eq. 3.1

where $\hat{p}_{g,Ak}$ is the frequency of variant $k$ at marker $g$ in the admixed population, $A$. Assuming that admixture is the sole determinant of allele frequencies in an admixed population formed from two source populations, for a given allele, $\hat{p}^A = m_1* p_1 + m_2* p_2$, where $m_1$ and $m_2$ are the proportionate contributions of source populations 1 and 2, and $p_1$ and $p_2$ are the respective allele frequencies in
each source. Here, to calculate $\hat{p}^A$, we used $m_{\text{African}} = 0.78$ and $m_{\text{European}} = 0.22$ based on previous genetic studies of admixture in African Americans (Parra et al., 1998, 2001b; Gravel, 2012; Bryc et al., 2015; Baharian et al., 2016; Gross, 2018).

Ancestry estimation

We used the model-based clustering method implemented in the program STRUCTURE to estimate individual ancestry (Pritchard et al., 2000). This program allows the user to set the number of clusters ($K$) into which it places each genotype or trait score of each individual. Every individual can therefore have membership in each cluster, with membership coefficients across all clusters summing to 100%. In a previous analysis of genetic data (Tishkoff et al., 2009), individuals from European and African source populations had approximately 100% membership in each of two distinctive genetic clusters, and African American individuals had variable membership in both clusters.

For each dataset, we ran STRUCTURE five times at $K = 2$ clusters using a burnin length of 25,000 and 15,000 MCMC repetitions. The analyses were unsupervised, meaning that prior information about the geographic origin of the samples was not used to assist in ancestry estimation. We also set the “advanced” options in STRUCTURE to print the 90% credible regions for individual ancestry estimates (Bayesian analogue to confidence intervals). STRUCTURE results were displayed using barplots.
Genetic data pruning

Anticipating our results, we found that the dichotomized and nondichotomized dental data had considerably lower total FI than the SNP and STR data. Given the relatively high FI for the SNP data and the fact that the Bayesian approach implemented in STRUCTURE is slow when applied to large datasets, we filtered the SNP dataset to loci with FI above 2.5, resulting in a set of 7,392 loci.

Reduced-FI genetic datasets

To determine if the different levels of FI in the dental and genetic datasets affected ancestry estimates, we extracted random subsets of genetic markers with similar levels of per-marker and total FI to the two forms of the dental dataset. For these “reduced-FI” analyses, for the SNP data, we returned to the full set of 992,601 markers. We subsetted the full SNP dataset to create a new SNP dataset with similar per-marker and total FI as the dichotomized dental data. We then subsetted the full STR dataset to create a new STR dataset with similar per-marker and total FI as the nondichotomized dental data. Our constraints were that (a) the new genetic datasets had 53 loci, the number of dental traits, and (b) the per-marker and total FI of the new genetic datasets were as close as possible to the per-marker and total FI of the dental data, given the first constraint.
The specific steps in this subsetting process were as follows. We first selected all genetic loci that had similar FI to each of the dental morphology markers (FI within 0.055 ensured that we selected one or more genetic loci for each dental trait). Next, we randomly selected one of the genetic loci that had been matched to each dental marker. This process resulted in a SNP dataset with 53 SNPs with per-marker and total FI similar to that of the dichotomized dental morphological data (given the above constraints), and it resulted in an STR dataset with 53 STRs and similar per-marker and total FI as the nondichotomized dental morphological data. We then re-ran STRUCTURE on these reduced-FI genetic datasets.

**Dichotomized vs. nondichotomized dental morphology**

The Wilcoxon Signed Rank Test is a nonparametric test used for paired data to test the null hypothesis that a randomly selected value from one sample is equally likely to be less than as it is greater than a randomly selected value from a second sample. This test is appropriate for proportion data and other data types with truncated distributions (e.g., at 0 and 1) (Klimentidis and Shriver, 2009; Yang et al., 2016; Healy et al., 2017). We used this test to determine whether the FI distributions from the dichotomized and nondichotomized dental morphological data were the same, and again to test whether the ancestry distributions produced by these two data forms were the same.
Results

Figure 3.1 shows histograms of FI for the datasets. All datasets have a high proportion of relatively uninformative markers (FI ≈ 0). Table 3.1 shows that the mean FI per marker was lowest in the SNP dataset, highest in the STR dataset, and intermediate for the two forms of the dental data. However, the SNP data contain thousands of markers with more information than the most informative dichotomized dental trait, and the total FI was more than 20,000 times greater. Similarly, the STRs had higher total FI than the nondichotomized dental morphological data and contained markers with higher FI than the most informative dental trait, though the differences between these datasets is less stark. Total FI was significantly higher for the nondichotomized data than for the dichotomized data (Wilcoxon Signed Rank Test, \( p < 2e^{-10} \)).
Figure 3. 1. Fisher Information. Y-axis values are scaled according to number of markers in the dataset

Figure 3.2 and Table 3.2 show the STRUCTURE results. In the SNP dataset, African individuals uniformly have nearly 100% membership in one of the inferred clusters, hereafter referred to as Cluster 1, and 0% membership in the other cluster, hereafter referred to as Cluster 2. Specifically, the mean membership of African individuals in Cluster 1 is 97.0% (Table 3.2), and the range of estimates is narrow (94.3–98.2%). The pattern is reversed for European individuals.
Table 3.1. Fisher Information* for full datasets and reduced-FI genetic datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Fisher Information</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Mean</td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>SNPs</td>
<td>225460.54</td>
<td>0.36</td>
<td>0.00</td>
<td>5.42</td>
</tr>
<tr>
<td>STRs</td>
<td>487.74</td>
<td>0.76</td>
<td>0.06</td>
<td>3.40</td>
</tr>
<tr>
<td>Dichotomized dental data</td>
<td>23.89</td>
<td>0.42</td>
<td>0.01</td>
<td>1.40</td>
</tr>
<tr>
<td>Non-dichotomized dental data</td>
<td>32.68</td>
<td>0.53</td>
<td>0.00</td>
<td>1.97</td>
</tr>
<tr>
<td>Reduced-FI SNP subset</td>
<td>17.64</td>
<td>0.31</td>
<td>0.00</td>
<td>1.69</td>
</tr>
<tr>
<td>Reduced-FI STR subset</td>
<td>32.78</td>
<td>0.52</td>
<td>0.06</td>
<td>1.92</td>
</tr>
</tbody>
</table>

* based on $m_{\text{African}} = 0.78$; $m_{\text{European}} = 0.22$

For African American individuals, membership in Cluster 1 varies between 58.3% and 91.7%, with a mean of 76.4%. In contrast with this wide range of Cluster 1 membership estimates, the error in individual estimates, as captured by the 90% credible regions, is, on average, less than 1% for each individual. In other words, African American individuals vary substantially in the amount of Cluster 1 membership.

The results are similar for the STR dataset. All African individuals have nearly 100% membership in Cluster 1 and the range of estimates is narrow (91.1–99.6%). Again, the pattern is reversed for European individuals, who uniformly have nearly 100% membership in Cluster 2. For African Americans, Cluster 1 membership ranges broadly from 45.3% to 98.4%, with a mean of 80.4%. The 90% credible regions for individual ancestry estimates are high compared to the SNP data, but they are still relatively narrow, on average, approximately 7% on either side of the estimate. Again, this result indicates that the level of Cluster 1 membership varies among African American individuals.
The clustering is less discrete for the dichotomized dental morphological data. Most African and European individuals shared appreciable membership in both genetic clusters. Nonetheless, on average, African individuals had higher membership in Cluster 1 than they did in Cluster 2, and European individuals had higher membership in Cluster 2 than in Cluster 1. For the dichotomized data, the mean membership of African individuals in Cluster 1 is 71.3% (range 9.9–95.6%) (Table 3.2). For the nondichotomized data, the mean membership of African individuals in Cluster 1, at 75.7% (range 9.2–97.0%), is significantly higher than for the dichotomized data (Wilcoxon Signed Rank tests, p < 0.05).

The credible regions for individual ancestry estimates are much broader than they were for the genetic datasets. For each individual in both forms of the dental morphological data, the credible regions typically extend 25% on either side of the Cluster 1 estimate. These broad error ranges overlap in the membership estimates for most individuals, implying the dental data are unable to capture interindividual differences in African (and European) ancestry among African Americans.
Figure 3.2. STRUCTURE results for all datasets. Each plot is comprised of a series of thin vertical bars, each representing an individual. Each bar is partitioned into two colored segments representing the individual's fractional membership in the two inferred clusters (Rosenberg et al., 2002). Samples are separated by black vertical lines, and individuals within each sample are ordered from lowest to highest proportion of Cluster 1 membership.
Reduced-FI genetic datasets

The last two rows of Table 3.1 show the FI for the genetic datasets that were subsetted to have similar FI to the two forms of the dental data. Table 3.3 shows the correlations in Cluster 1 membership between the reduced-FI genetic and full genetic datasets. Figure 3.3 shows STRUCTURE results for the subsetted datasets. For the reduced-FI SNP dataset, African and European individuals still fall almost exclusively into distinctive clusters. However, there is considerably more variation in membership in the two clusters among African and European individuals, and the 90% credible regions for individual estimates are broader. For African individuals, for example, membership in Cluster 1 varies from 42.8% to 99.5% (mean = 96.3%), and the average 90% credible regions for individual estimates range from 12.8% below to 3.7% above the estimate. For African Americans, the mean membership in Cluster 1 is higher than it was for the full SNP dataset, at 85.6%, and the range of Cluster 1 membership estimates are much broader (22.4–99.5%).

For the reduced-FI STR dataset, as was the case with the nondichotomized dental data, most African and European individuals share appreciable membership in both genetic clusters, though African individuals had higher membership in Cluster 1 than they did in Cluster 2, and European individuals had higher membership in Cluster 2 than they did in Cluster 1. For African Americans, the mean membership in Cluster 1, at 68.8%, was 11.6% lower than it was for the full STR dataset, and the range of Cluster 1 estimates
was broader (23.4–93.7%). Furthermore, the credible regions broadened appreciably compared to the full STR dataset, to, on average, approximately 20% on either side of the Cluster 1 estimates. As with the dental data, these ranges are too broad to capture interindividual differences in African (and European) ancestry among African Americans that were clearly captured in the full STR dataset.

Nonetheless, the reduced-FI STRs still perform better than the nondichotomized dental data in some respects. For example, looking at the range of Cluster 1 membership in Table 3.2 (corresponding to African ancestry), the highest membership for Europeans in the nondichotomized dental data is 95.0%, whereas for the reduced-FI STRs, the highest membership for Europeans is 23.4%.

**Table 3.2. Cluster 1 membership**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Sample</th>
<th>Mean Cluster 1 membership (st dev)</th>
<th>Cluster 1 Range</th>
<th>Mean Cluster 2 membership (st dev)</th>
<th>Cluster 2 Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNPs</td>
<td>African</td>
<td>0.970 (0.006)</td>
<td>0.943 - 0.982</td>
<td>0.030 (0.006)</td>
<td>0.017 - 0.055</td>
</tr>
<tr>
<td></td>
<td>European</td>
<td>0.013 (0.010)</td>
<td>0.000 - 0.039</td>
<td>0.987 (0.010)</td>
<td>0.951 - 0.998</td>
</tr>
<tr>
<td></td>
<td>African</td>
<td>0.764 (0.077)</td>
<td>0.583 - 0.917</td>
<td>0.236 (0.077)</td>
<td>0.081 - 0.412</td>
</tr>
<tr>
<td></td>
<td>European</td>
<td>0.021 (0.015)</td>
<td>0.003 - 0.056</td>
<td>0.979 (0.015)</td>
<td>0.953 - 0.998</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>0.804 (0.101)</td>
<td>0.453 - 0.984</td>
<td>0.196 (0.101)</td>
<td>0.009 - 0.561</td>
</tr>
<tr>
<td>STRs</td>
<td>African</td>
<td>0.984 (0.015)</td>
<td>0.911 - 0.996</td>
<td>0.016 (0.015)</td>
<td>0.002 - 0.112</td>
</tr>
<tr>
<td></td>
<td>European</td>
<td>0.021 (0.015)</td>
<td>0.003 - 0.056</td>
<td>0.979 (0.015)</td>
<td>0.953 - 0.998</td>
</tr>
<tr>
<td></td>
<td>African</td>
<td>0.804 (0.101)</td>
<td>0.453 - 0.984</td>
<td>0.196 (0.101)</td>
<td>0.009 - 0.561</td>
</tr>
<tr>
<td></td>
<td>European</td>
<td>0.021 (0.015)</td>
<td>0.003 - 0.056</td>
<td>0.979 (0.015)</td>
<td>0.953 - 0.998</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>0.804 (0.101)</td>
<td>0.453 - 0.984</td>
<td>0.196 (0.101)</td>
<td>0.009 - 0.561</td>
</tr>
<tr>
<td>Dichotomized dental</td>
<td>African</td>
<td>0.713 (0.200)</td>
<td>0.099 - 0.956</td>
<td>0.287 (0.200)</td>
<td>0.044 - 0.901</td>
</tr>
<tr>
<td></td>
<td>European</td>
<td>0.284 (0.202)</td>
<td>0.043 - 0.927</td>
<td>0.716 (0.201)</td>
<td>0.073 - 0.957</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>0.595 (0.218)</td>
<td>0.063 - 0.949</td>
<td>0.405 (0.218)</td>
<td>0.051 - 0.937</td>
</tr>
<tr>
<td>Non-dichotomized dental</td>
<td>African</td>
<td>0.757 (0.196)</td>
<td>0.092 - 0.970</td>
<td>0.243 (0.196)</td>
<td>0.030 - 0.908</td>
</tr>
<tr>
<td></td>
<td>European</td>
<td>0.253 (0.207)</td>
<td>0.029 - 0.950</td>
<td>0.747 (0.207)</td>
<td>0.050 - 0.971</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>0.572 (0.249)</td>
<td>0.043 - 0.966</td>
<td>0.427 (0.249)</td>
<td>0.034 - 0.957</td>
</tr>
<tr>
<td>Reduced-FI SNP subset</td>
<td>African</td>
<td>0.963 (0.071)</td>
<td>0.428 - 0.995</td>
<td>0.037 (0.071)</td>
<td>0.005 - 0.572</td>
</tr>
<tr>
<td></td>
<td>European</td>
<td>0.033 (0.043)</td>
<td>0.005 - 0.263</td>
<td>0.967 (0.043)</td>
<td>0.737 - 0.995</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>0.856 (0.196)</td>
<td>0.224 - 0.995</td>
<td>0.144 (0.196)</td>
<td>0.005 - 0.776</td>
</tr>
<tr>
<td>Reduced-FI STR subset</td>
<td>African</td>
<td>0.683 (0.182)</td>
<td>0.231 - 0.906</td>
<td>0.317 (0.182)</td>
<td>0.094 - 0.769</td>
</tr>
<tr>
<td></td>
<td>European</td>
<td>0.134 (0.043)</td>
<td>0.066 - 0.243</td>
<td>0.866 (0.043)</td>
<td>0.757 - 0.934</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>0.625 (0.162)</td>
<td>0.247 - 0.877</td>
<td>0.375 (0.162)</td>
<td>0.123 - 0.753</td>
</tr>
</tbody>
</table>

*averaged across individuals
Table 3.3. Correlation in Cluster 1 membership for full and reduced-FI genetic datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Sample</th>
<th>Pearson's R</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced-FI SNP subset</td>
<td>African</td>
<td>0.798</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td>European</td>
<td>0.794</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td>African American</td>
<td>0.887</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Reduced-FI STR subset</td>
<td>African</td>
<td>0.901</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td>European</td>
<td>0.965</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td>African American</td>
<td>0.984</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

Figure 3.3. STRUCTURE output for reduced-FI genetic datasets.

Discussion and conclusions

We found that the mean per-marker FI was broadly similar in the dental and genetic datasets. However, because the number of genetic markers greatly exceeded the number of dental traits, the genetic datasets contain substantially higher total FI than both forms of the dental data. For these high-FI genetic
datasets, we replicated the results of previous studies in showing that: (a) African and European individuals fall almost exclusively into distinct genetic clusters (Rosenberg et al., 2002; Jakobsson et al., 2008; Li et al., 2008), and (b) average population-level African ancestry in African Americans ranges between 75% and 85% (Parra et al., 1998, 2001b; Bryc et al., 2010; Gravel, 2012; Baharian et al., 2016; Gross, 2018). These patterns are in contrast to those of the dental morphological dataset, both in dichotomized and nondichotomized form, for which the African and European samples did not form distinct clusters, and African Americans had low mean African ancestry relative to these previous studies.

We noted above that the genetic and dental morphological samples are from different individuals in different geographic locations, which may have contributed to the difference in ancestry estimates among African Americans in the genetic and dental datasets. However, given that (a) population-level African ancestry estimates for the genetic data only vary between roughly 75–85% across regions of the United States, and (b) the African and European dental samples did not form distinct clusters in the STRUCTURE analysis, we conclude that the discrepant results for the genetic and dental datasets are not because of differences in sampling locations.

To test whether the better performance of the genetic data was due strictly to the larger number of genetic markers, we subsetted the SNP and STR datasets to match the dental data in number of markers, per-marker FI, and total
Analyses of these reduced-FI genetic datasets confirm that the discrepant results between the dental and genetic data are due in part to relatively low FI in the dental data. However, while the reduced-FI SNPs performed markedly worse in STRUCTURE than the initial SNP dataset, the performance was still much better than for the dichotomized dental data.

Similarly, we found that the reduced-FI STRs performed worse than the full STR dataset but better than the nondichotomized dental data. The relatively poor performance of the dental data in part reflects the relatively high amount of missing data relative to the SNP and STR datasets. SNPs had the lowest amount of missing data, followed by the STRs, and the dental data had the highest amount. STRUCTURE is unable to assign a missing marker in an individual to a source population. Therefore, individuals with high proportions of missing data will have high membership in all clusters.

To consider other possible contributors to the poorer performance of the dental data, it is important to revisit the assumptions that we made at the beginning of the study. The first assumption was that African American populations were formed through mating between sub-Saharan Africans and Europeans. Because other regional populations have contributed to African American populations, this assumption is not strictly met. Bryc et al. (2015), for example, found that, nationwide, about 5% of African Americans carry at least 2% Native American ancestry. The proportion of African Americans with Native American ancestry is higher in the Southwest than in other regions, though, even
here, the average per-individual Native American ancestry is less than 2%. For the HapMap SNP data used in this study, we were unable to estimate Native American ancestry because the dataset lacked a Native American sample. For the STR data, using Native American data from Tishkoff et al. (2009) and running STRUCTURE at $K = 3$, Gross (2018) estimated an average Native American contribution of less than 1%. These low proportions would have had no appreciable effect on the estimation of FI or African and European ancestry in our study.

The second assumption was that admixture is the sole determinant of allele and trait frequencies in African Americans, that is, post-admixture drift and selection played no role. Under this assumption, the allele and trait frequencies in the admixed African American population are a function of the allele frequencies in the African and European source populations and their proportionate contributions to the African American population. These ancestral allele frequencies are, in turn, a key determinant of FI (see Equation 3.1). If these frequencies were affected by post-admixture drift or selection, the utility of FI as a means of assessing the informativeness of the data would be limited. With respect to drift, founder effects may have shaped allele frequencies in the European and African populations that migrated to the United States over the past 500 years. For this reason, it is important to note that the precise source populations are unknown. Instead admixture studies use proxy populations from
Africa and Europe and the United States (e.g., European Americans). These factors no doubt lead to error in the estimation of FI and ancestry in African American and other admixed populations throughout the Americas.

Homoplasy may also have affected allele frequencies in source and admixed populations. Because of a high rate of mutation (Brinkmann et al., 1998), homoplasy is common in STRs, as it may be in dental morphological traits. The comparative mutation rate for SNPs is several orders of magnitude lower than for STRs (Nachman and Crowell, 2000), leading to comparatively low levels of homoplasy. This fact may explain the strong performance of the reduced-FI SNP dataset relative to the reduced-FI STR dataset.

The third assumption was that dental traits are reflective of underlying genetic variation. While there is a consensus among anthropologists that dental morphology is determined predominantly by genetic factors, little is known about the genetic architecture of morphological traits. One exception relates to incisor shoveling, one of the traits used in this study. Kimura et al. (2009) identified an association between alleles of the ectodysplasin A receptor (EDAR) gene and the degree of incisor shoveling in a sample of Japanese individuals. The effect of the alleles was additive and explained about 19% of the variation in the degree of shoveling. Similarly, Tan et al. (2014) found an association between EDAR alleles and the degree of shoveling in the Uyghurs of Xinjiang, western China. Future studies of the genetic basis of dental morphology will help to identify traits that are well suited for studying the evolutionary processes that have molded
diversity within and among human populations. Such studies may also help to optimize the breakpoints used to convert nondichotomized traits to dichotomized traits.

Despite the low FI of the dental data, we found that (a) African and European individuals had consistently higher membership in one cluster than the other, and (b) African American individuals had consistently higher membership in the same cluster in which African individuals had high membership. These results imply that dental data can distinguish African from non-African populations. Given the success of dental data in distinguishing both global and regional samples from one another (e.g., Hanihara, 2008; Edgar, 2015; Irish, 2015), this result is unsurprising. However, a novel finding of our study is that, in cases where source populations are relatively deeply diverged, dental morphology can potentially identify admixed populations that formed from those sources. Further analyses of the type presented here would be required to determine if dental morphology can detect admixture between more recently diverged populations, for example, Europeans and Native Americans (see also Edgar, 2017).

We found that the credible regions for individual ancestry estimates for the full genetic datasets were narrow, implying that genetic data have the potential to detect interindividual differences in African and European ancestry in African American populations. In contrast, the credible regions were broad for dental morphology and reduced-FI genetic datasets, indicating that they are not ideally
suited to detecting differences in admixture proportions between individuals within admixed populations. However, such datasets are useful in the investigation of interpopulation differences, and the investigation of simple admixture models, particularly in cases where DNA sequences are unavailable.

Finally, we found that nondichotomized dental data have significantly higher FI than dichotomized data, and that nondichotomized data produce STRUCTURE results that are more similar to our expectation given results from genetic analyses (Parra et al., 1998, 2001b; Gravel, 2012; Bryc et al., 2015; Baharian et al., 2016). Nichol and Turner (1986) state that a third of scoring differences because of intra- and inter-observer error occur at the breakpoint. Additionally, the reduction in overall trait variation that is caused by dichotomization may mask information about genetic relationships within and between populations (Mayhall, 1999; Harris, 2008). Our finding that nondichotomized dental morphological data have significantly higher FI and perform better in ancestry estimation supports this view. This finding indicates that nondichotomized dental data are better suited for ancestry estimation in admixed populations, and may be better suited in general for detecting differences among human populations.
CHAPTER 4. VARIATION IN AFRICAN ANCESTRY AMONG AFRICAN AMERICANS ACROSS SPACE AND TIME

Abstract

Objectives:

Our goals in this study are to 1) evaluate differences in African ancestry in African Americans across time and geographic location, 2) determine the possible causes of the differences in African ancestry, 3) further define the power of dental morphology to study the evolutionary process.

Methods:

Our dataset consisted of nondichotomized dental morphological data from 30 sub-Saharan Africans, 214 Europeans and European Americans, and 486 African Americans from two time periods, and six locations across the US. Our analyses included: 1) estimating individual ancestry proportions using PCA and the program STRUCTURE to produce individual ancestry distributions for each location, 2) testing the hypothesis that individuals differed significantly from the sample mean ancestry, and 3) testing whether the distributions of individual ancestry were different across geographic locations and time periods.
Results:

We found that distributions of individual ancestry vary significantly between two samples of African Americans from six geographic locations of the United States. Individual African ancestry within each sampling location varied widely from 6.6% to 94.4%. On average, 68.6% of African Americans differed significantly from the mean African ancestry within their sampling location.

Mean European ancestry in the African Americans samples ranged from 36.9% to 47.9%. PCA analysis showed that African American trait frequencies were intermediate between African and European trait frequencies for all samples along the axis joining the African and European samples. African American samples also separated along an orthogonal axis.

Discussion:

Although limited by missing data and number of observations, dental morphological data are a useful resource for studying the admixture process of present and past populations. I identified structure due to admixture and drift in the African American population across space and time.
Introduction

Admixed populations form when two previously separated populations come together and exchange mates. The African American population results primarily from the exchange of mates between Africans and Europeans over the past few centuries during which time both Europeans and immigrant Africans contributed to the population (Gross, 2018). These contributions from Europeans and Africans has likely varied over the course of African American history. For example, while ending legal slave importation in 1808 did not fully stop the importation of new African slaves, it may have led to a decrease in number of individuals brought to the US from Africa, and the passage of anti-miscegenation laws prevented European Americans from taking legal African American wives, thereby perhaps decreasing the unions between them, it did not stop mating between the two groups (General Assembly of Maryland, 1664; Curtin, 1969; Smith, 1973).

While the exact dynamics of African American admixture history is unknown, the social history of this group is well-documented. There is a dearth of historical data specific to the timing and rate of admixture over the past 400 years; however, the historical record does provide information that can be used to constrain the parameters of admixture models. For example, we know that the ancestors of African Americans came from various locations within Africa, including areas where admixture with people from other regions occurred, such as regions that border the Mediterranean. However, the vast majority of
immigrants both during and after the slave came from Western sub-Saharan Africa (Curtin, 1969; Voyages Database, 2009; US Census Bureau, 2010). We also know that the first importation event to North America involved 20 slaves in Virginia in 1619 (Curtin, 1969). After this initial event, slave importation remained low until the beginning of the 18th century (see Figure 4.1), after which it continued until 1860, even after importation became illegal in 1808 (Curtin, 1969; Smith, 1973). The passage of anti-miscegenation laws as early as 1664 suggest that African American-European mating occurred early in US history (General Assembly of Maryland, 1664). Additionally, reports made by slaves discuss forced mating between slaves and slave owners throughout African American history (Federal Writers Project, 2001). More recently, the 2010 Census reports 314,400 immigrants from Northern African and 2,847,199 new immigrants from sub-Saharan Africa (US Census Bureau, 2010). Based on this information, it is likely that admixture between Europeans and African Americans began as early as 12 generations before the present and that it occurred continuously afterwards. It is also likely that the rate of European gene flow into the African American population varied dramatically over the past 400 years as a result of important historical events such as the US Civil War, the passage of anti-miscegenation laws (pre- and post-Civil War), the Great Migration, and the passage of Civil Rights legislation.

Some of these important historical events may have had asymmetrical effects on the rate of European gene flow into the African American population in
different regions of the US. One such event is the Great Migration, which was a period from 1916 to 1970 during which six million African Americans moved from the rural South to escape Jim Crow legislation, racial violence, and inferior educational opportunities, and for the economic opportunity of the industrial, urban North (Mandle, 1978; Ransom & Sutch, 1977; Tolnay, 2003; Tolnay & Beck, 1995). This event dramatically reduced both the absolute number of African Americans living in the South and the proportion of the population in the South that they comprised, and increased the number and proportion of African Americans living in the North.

Gross (2018) showed that samples of African Americans in different geographic locations in the US have different individual ancestry distributions. Bryc et al. (2015) and Baharian et al (2016) showed that the average African ancestry of African Americans varies by state. Bryc et al. (2015) showed that African Americans in the US Northeast had higher average European ancestry than those in the US South. The cause of this difference is unclear. One possible explanation is a difference history of race relations over the entire course of African American history resulting in consistently higher rates of European admixture in the north. Alternatively, there may have been an increase in admixture during and after the mass relocation of African Americans during the Great Migration. The difference may also be due to higher European ancestry among the African Americans who migrated north than in the general population.
of African Americans in the South, which may be a result of differential opportunity among African American individuals based on ancestry.

It is possible that the variation in ancestry among African Americans is structured not only across geographic space, but also across time. Admixture studies on living people, such as those typically performed on genetic data, consider individuals in their current geographic location. However, the ancestors of contemporary African Americans may have lived in different locations prior to events like the Great Migration. Alternative data types, such as cranial and dental morphological traits are available for African Americans, European Americans, and Africans, and are situated in both geographic space and time. Admixture studies on these data have the potential to provide information about the pattern of variation in African Americans in the past in a way that genetic data from contemporary people cannot. Analysis of dental morphology traits from historical cemetery collections allows us to estimate ancestry in individuals situated in time and space. Analyzing dental morphology traits from contemporary burials allows us to compare the ancestry of recently living individuals to those from historical collections.

Our goals in this study are to: 1) determine whether we can detect differences in African ancestry in African Americans from two different time periods and various geographic locations; 2) determine the possible causes of the differences we observe, and; 3) further define the capabilities of dental morphology to elucidate the evolutionary process.
Methods

Data

We used 63 ordinally-graded dental morphology traits (hereafter referred to as “nondichotomized”) from contemporary and 19th and early 20th Century African Americans, West Africans, Europeans, and European Americans. All data were collected and scored as graded by Heather JH Edgar according to Turner II et al. (1991). The initial dataset consisted of 15 samples of African Americans: four contemporary samples, and 11 historic samples from the 19th and 20th Century. To minimize error in our ancestry estimates, we reduced the dataset to traits scored in at least 50 individuals per population (i.e., African, European, and African American populations), and individuals with at least 80% of traits scored. This resulted in a set of 53 morphology traits from 515 African Americans from contemporary samples from four locations and 19th and early 20th Century samples from nine locations, 214 Europeans from 13 locations, and 30 West Africans from 14 locations. We further reduced the dataset to African American samples that retained a minimum of 20 individuals, resulting in a sample of 486 African Americans including four contemporary samples and two 19th/early 20th Century samples, which we will call “Historic” (see Table 4.1 below).

Table 4.1 summarizes the African American samples. Contemporary African American samples were from orthodontic collections, and race was assigned by the orthodontist who assembled the collections. The historic African American samples include individuals interred in cemeteries reserved for African
Americans. We use two historic samples in our analyses: 1) the Freedman’s Cemetery sample, which includes burials from 1867 to 1907, and 2) the Gullah sample which includes burials from 1913 to 1954. The four contemporary samples include individuals born in the mid to late 20th century.

Table 4.1. African American samples used in analyses.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Region</th>
<th>Location</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contemporary</td>
<td>New York University (NYU)</td>
<td>New York</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>University of Tennessee (UT)</td>
<td>Tennessee</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>University of Washington (UW)</td>
<td>Washington</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>University of Southern California (USC)</td>
<td>Southern California</td>
<td>40</td>
</tr>
<tr>
<td>Historic</td>
<td>Freedman’s Cemetery</td>
<td>Texas</td>
<td>254</td>
</tr>
<tr>
<td></td>
<td>Gullah</td>
<td>South Carolina/Georgia</td>
<td>26</td>
</tr>
</tbody>
</table>

Ancestry Estimation

We used the model-based clustering algorithm implemented in the program STRUCTURE to estimate individual ancestry (Pritchard et al., 2000) for the full datasets and subsetted datasets. Average Native American ancestry is low in African Americans (Parra et al., 1998; Bryc et al., 2015; Baharian et al., 2016; Gross, 2018), therefore we assumed a dihybrid admixture model. We ran STRUCTURE five times at K=2 using a burnin length of 25,000 steps and 15,000 MCMC repetitions. We did not supervise our analyses, meaning we did not use prior information about the geographic origins of the samples to assist in ancestry estimation. We used the “advanced” tab in STRUCTURE to have the program
print the 90% credible regions (Bayesian analogue to confidence intervals) for
the individual ancestry estimates to determine the precision of the estimates.

**Variation Within Samples**

We used the 90% credible regions generated by STRUCTURE to
determine whether African American individuals differed from the sample mean
African ancestry. We did this to determine whether variation in individual ancestry
estimates could be detected using dental morphological traits. This determination
also constitutes a test of a one-time admixture model in which Africans and
Europeans came together at a single time in history to form the African American
population.

**Variation Among Samples**

We used two approaches to examine variation among African American
samples. First, we used a Wilcoxon Rank Sum Test to determine whether the
distributions of individual cluster membership given by STRUCTURE varied
significantly among the African American samples. This is a nonparametric test
for unpaired data used to determine whether a randomly selected value from one
distribution is greater than or less than a randomly selected value from another
distribution. The null hypothesis for this test is that the randomly selected value
from one distribution is equally likely to be greater than as it is to be less than the
randomly selected value from the second distribution. This test has been used in
previous studies for proportion data and other data types with distributions truncated at 0 and 1 (Klimentidis and Shriver, 2009; Yang et al., 2016; Healy et al., 2017; Gross and Edgar, 2019). Our second approach was to use principal component analysis (PCA) to summarize the major axes of trait variation among all of our samples. PCA analysis was conducted using the ade4 package (Dray and Dufour, 2007) in R (R Core Team, 2018). Missing data were replaced with zeros.

Results

Ancestry estimation

We used STRUCTURE at $K=2$ to estimate individual ancestry in our dataset, to determine the proportion of European and African ancestry in our African American samples. Table 2.4 summarizes the STRUCTURE results for all samples. The African sample had higher mean Cluster 1 membership (77.8%) than Cluster 2 membership (22.2%), and the European sample had higher mean Cluster 2 (74.6%) than Cluster 1 membership (25.4%). We infer from this that Cluster 1 roughly corresponds to African ancestry. Although mean Cluster 1 membership was relatively high for the African sample, individual estimates varied broadly, with the lowest estimate being 33.9% and the highest being 96.1%. The European range was even broader, with a low estimate of 4.1% and a high of 93.6%. All African American samples had higher mean Cluster 1 than mean Cluster 2 membership, and mean Cluster 1 membership for all African
American samples were intermediate between those for African and European samples, ranging from 52.1% in the UW sample to 63.1% in the NYU sample. As with the African and European individual Cluster 1 membership, individual estimates for African Americans in the dataset ranged broadly from 6.6% to 94.4%.

Table 4.2. STRUCTURE results for Cluster 1 membership (Cluster 2 = 1 - Cluster 1 membership).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Sample</th>
<th>Mean Cluster 1 fractional membership (st dev)</th>
<th>Cluster 1 range</th>
</tr>
</thead>
<tbody>
<tr>
<td>African</td>
<td></td>
<td>0.778 (0.168)</td>
<td>0.339 - 0.961</td>
</tr>
<tr>
<td>European</td>
<td></td>
<td>0.254 (0.191)</td>
<td>0.041 - 0.936</td>
</tr>
<tr>
<td>Contemporary</td>
<td>NYU</td>
<td>0.631 (0.188)</td>
<td>0.192 - 0.943</td>
</tr>
<tr>
<td></td>
<td>UT</td>
<td>0.524 (0.236)</td>
<td>0.066 - 0.944</td>
</tr>
<tr>
<td></td>
<td>UW</td>
<td>0.521 (0.232)</td>
<td>0.129 - 0.931</td>
</tr>
<tr>
<td></td>
<td>USC</td>
<td>0.602 (0.209)</td>
<td>0.140 - 0.942</td>
</tr>
<tr>
<td>Historic</td>
<td>Freedman’s Cemetery</td>
<td>0.578 (0.239)</td>
<td>0.091 - 0.943</td>
</tr>
<tr>
<td></td>
<td>Gullah</td>
<td>0.574 (0.208)</td>
<td>0.095 - 0.931</td>
</tr>
</tbody>
</table>

Variation within samples

The 90% credible regions for individual cluster membership estimates from STRUCTURE are fairly broad, with an average of 20% above and 20% below the estimate. Filtering the data to individuals with 20% or less missing data reduced the credible regions relative to those Gross and Edgar (2019) found of 25% on either side of the estimate, however, these credible regions are still broad relative to those produced by genetic data (Gross and Edgar, 2019). Table 3.4 shows the number and proportion of individuals in each African American
sample that differ from the sample mean Cluster 1 membership. A high proportion of individuals from each sample have credible regions that encompass the mean Cluster 1 membership. Sample sizes vary among samples, but the Gullah have the highest proportion of individuals with mean membership (80.8%), and the Freedman’s Cemetery sample has the lowest proportion (61.0%).

**Table 4.3. Proportions of individuals who differ from sample mean African ancestry**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Sample</th>
<th>n</th>
<th>Number of individuals with mean Cluster 1 membership (proportion of sample)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contemporary</td>
<td>NYU</td>
<td>54</td>
<td>41 (0.759)</td>
</tr>
<tr>
<td></td>
<td>UT</td>
<td>90</td>
<td>61 (0.678)</td>
</tr>
<tr>
<td></td>
<td>UW</td>
<td>22</td>
<td>15 (0.682)</td>
</tr>
<tr>
<td></td>
<td>USC</td>
<td>40</td>
<td>28 (0.700)</td>
</tr>
<tr>
<td>Historic</td>
<td>Freedman’s Cemetery</td>
<td>254</td>
<td>155 (0.610)</td>
</tr>
<tr>
<td></td>
<td>Gullah</td>
<td>26</td>
<td>21 (0.808)</td>
</tr>
</tbody>
</table>

We used the STRUCTURE output to assess individual African ancestry as proxied by Cluster 1 membership in the African American samples. Figure 4.1 shows the distribution of Cluster 1 membership for all African American samples along with the sample sizes. These distributions show that individual Cluster 1 membership varies broadly within each sample from less than 20%, and as low as 6.6% in the UT sample, to over 90%. All samples had individuals in the highest membership bin (90-100% Cluster 1 membership). The UT, Freedman’s
Cemetery, Gullah samples had individuals in the lowest membership bin (0% to 10% Cluster 1 membership).

**Figure 4.1.** Distributions of Cluster 1 membership. Top row: contemporary samples, bottom row: historic samples.

**Variation among samples**

To assess variation in ancestry across space and time, we evaluated variation among our samples African Americans in two ways, 1) comparing distributions of Cluster 1 membership, and 2) using PCA. While the distributions of individual Cluster 1 membership appear to be different shapes for each sample, Wilcoxon Rank Sum Tests (Table 4.4) show that the only significant differences are between the NYU and UT samples ($p = 0.008$), and NYU and UW
The differences between all other pairs of samples are statistically indistinguishable.

Table 4.4. Pairwise Wilcoxon Rank Sum Tests for African American individual Cluster 1 membership distributions.

<table>
<thead>
<tr>
<th>sample 1</th>
<th>sample 2</th>
<th>W</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>NYU</td>
<td>UT</td>
<td>3078</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>UW</td>
<td>768.5</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>USC</td>
<td>1151</td>
<td>0.590</td>
</tr>
<tr>
<td></td>
<td>Freedman’s Cemetery</td>
<td>7616.5</td>
<td>0.202</td>
</tr>
<tr>
<td></td>
<td>Gullah</td>
<td>805.5</td>
<td>0.290</td>
</tr>
<tr>
<td>UT</td>
<td>UW</td>
<td>1011</td>
<td>0.881</td>
</tr>
<tr>
<td></td>
<td>USC</td>
<td>1440</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>Freedman’s Cemetery</td>
<td>9934.5</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td>Gullah</td>
<td>992.5</td>
<td>0.241</td>
</tr>
<tr>
<td>UW</td>
<td>USC</td>
<td>345</td>
<td>0.164</td>
</tr>
<tr>
<td></td>
<td>Freedman’s Cemetery</td>
<td>2401</td>
<td>0.275</td>
</tr>
<tr>
<td></td>
<td>Gullah</td>
<td>243</td>
<td>0.379</td>
</tr>
<tr>
<td>USC</td>
<td>Freedman’s Cemetery</td>
<td>5280</td>
<td>0.690</td>
</tr>
<tr>
<td></td>
<td>Gullah</td>
<td>545.5</td>
<td>0.743</td>
</tr>
<tr>
<td>Freedman’s Cemetery</td>
<td>Gullah</td>
<td>3397.5</td>
<td>0.809</td>
</tr>
</tbody>
</table>

Figure 4.2 is a scatter plot of PC 1 and PC 2 for our full dataset of Africans, Europeans, and African American dental morphology. PC 1 explains 2.786%, and PC 2 explains 2.429% of the variation in trait frequencies in the dataset. Both PC 1 and PC 2 separate Africans and Europeans, and all African American samples, except the Freedman’s Cemetery sample, are intermediate between the Africans and Europeans. PC 1 groups the Freedman’s Cemetery sample with Africans, and PC 2 groups them with Europeans. As shown in the figure, while there is overlap between the African American samples, there
appear to be three clusters of African Americans: 1) a UT and Gullah cluster, 2) a USC, UW, and NYU cluster, and 3) a Freedman’s Cemetery cluster. As with the STRUCTURE results, the PCA shows a broad range of variation among individuals within each sample. As with the distributions, NYU and UT fall into different groups, supporting the results from the Wilcoxon tests.

Figure 4.2. PCA 1 and 2. PC 1 is on the x-axis, and PC 2 is on the y-axis. FC = Freedman’s Cemetery, Gu = Gullah, NYU = New York University, USC = University of Southern California, UT = University of Tennessee, UW = University of Washington.

Discussion

We had three primary goals in this study. The first was to evaluate the pattern of ancestry in African Americans across time and geographic location.
We began by estimating individual ancestry in each of our African American samples. Gross and Edgar (2019) showed that nondichotomized dental data are useful in distinguishing mean differences in ancestry between populations and identifying admixed populations, but that these data were unable to distinguish differences among individuals within populations. Delgado et al. (2018) came to the same conclusion using a sample of Hispanics from Colombia, South America. In their study, Delgado et al. found that dental data provide relatively good estimates of average population genetic ancestry, but less useful individual estimates. Left unclear by these two studies was whether dental morphological data could be used to distinguish ancestry differences between subgroups of admixed populations.

All samples of African Americans had mean cluster membership that was intermediate between the mean for Africans (77.8%) and the mean for Europeans (25.4%), ranging from 52.1% to 63.1%. This is consistent with the expectation for an admixed population, in which trait frequencies should be a linear combination of the source frequencies (Pfaff et al., 2004), and consistent with the findings of Gross and Edgar (2019). However, the mean Cluster 1 membership was relatively low compared to those found in genetic studies of African American ancestry (Parra et al., 1998, 2001b; Gravel, 2012; Bryc et al., 2015; Baharian et al., 2016), which range between 75% and 85%.

Because of the broadness of the credible regions, complex admixture models could not be tested. However, we were able to use the credible regions
to determine whether individuals in each sample differ significantly from the sample mean ancestry. Verdu and Rosenberg (2011) demonstrated that variance in ancestry decreases, and the distribution of individual ancestry becomes increasingly concentrated around the mean across generations after a one-time admixture event. Gross (2018) [NO_PRINTED_FORM] demonstrated that the variance in individual ancestry is low by 12 generations (the time depth of the African American population) after a one-time admixture event, and reaches zero by 20 generations. That same study concluded that a one-time admixture model did not fit the distribution of ancestry for two genetic datasets, and other studies of genetic data have found variation in ancestry among individuals within and between sampling locations (Tishkoff et al., 2009; Bryc et al., 2015). In our study, a high proportion of individuals in all African American samples were not significantly different from the sample mean ancestry. The discrepancy between our results and those of genetic studies suggests that the lack of power in the dental morphological data precludes testing specific admixture models. Instead, dental morphological data may be better suited to gross comparisons between samples.

Bryc et al. (2015) showed that, among the states from which our samples originate, South Carolina and Texas have the highest African ancestry. Based on the results of this study, we expected that the Gullah and the Freedman’s Cemetery samples would have the highest mean Cluster 1 membership among our samples. In contrast, we found the Gullah sample to have 57.4% Cluster 1
membership and Freedman’s Cemetery to have 57.8%, and the NYU sample has the highest membership.

The Gullah sample had lower Cluster 1 membership than expected given the results of genetic studies, which show high levels of African ancestry in the Gullah compared to other African American groups (Parra et al., 2001; Garvey et al., 2003; McLean et al., 2005), and historical information which indicates that this group was reproductively isolated from both Europeans and other African Americans (Cassidy, 1980; Opala and Timmons, 1987; Pollitzer, 1999). Mean Cluster 1 membership in the Gullah was 54.7%, and the PCA shows their trait frequencies to be nearly exactly intermediate between Africans and Europeans for PC 1 and PC 2. This may be due to the small sample size for the Gullah (n=26). Alternatively, this may result from using an incorrect source samples for the Gullah for comparison. The African ancestors of the Gullah can be traced back to rice-growing regions of Sierra Leone specifically (Opala and Timmons, 1987; Pollitzer, 1999; Parra et al., 2001; McLean et al., 2005). While the samples we use in our analyses may be reasonable for other African Americans whose ancestors are from West Africa more generally, these populations may not be appropriate for the Gullah.

Following other studies of admixed populations (Patterson et al., 2006; Ma and Amos, 2012; Healy et al., 2017), we used PCA to examine structure within our African Americans samples. In an admixed population, allele frequencies are intermediate between the frequencies of the source populations if admixture is
the only force of evolution at play (Pfaff et al., 2004; McLean et al., 2005; Patterson et al., 2006; Ma and Amos, 2012). Studies of dental morphological traits have recapitulated patterns of neutral genetic variation among human populations (Irish and Guatelli-Steinberg, 2003; Hanihara, 2008), indicating that these data are a potential proxy for neutral genetic variation. Therefore, PCA provides a way of visualizing the variation in trait frequencies in our samples. The PCA of our samples indeed showed most African American samples to fall between the African and European samples as is expected for an admixed population.

The STRUCTURE results revealed differences among samples of African Americans. We found that dental data can be used to distinguish between some subgroups in the African American population. The Wilcoxon Rank Sum Tests showed that the distributions of individual Cluster 1 membership for the NYU sample differed significantly from both the UT among the African American samples, and UW and UT have the lowest (52.1% and 52.4%, respectively). According to the Bryc et al. (2015) study, African Americans in New York indeed have higher African ancestry than African Americans in Washington, however, they show that African Americans in Tennessee have higher African ancestry than those in New York.

However, it is unclear from the STRUCTURE analysis alone whether these differences are due strictly to admixture or if other evolutionary processes, such as genetic drift, are at play. The second goal of this study was to better
understand the causes of variation in African American ancestry. Evidence that other evolutionary processes affected the African American samples is visible in the PCA plot itself. Trait frequencies in an admixed population are a linear combination of the frequencies in the sources if admixture has been the sole evolutionary process affecting the admixed population. Deviations from linearity could be due to the use of incorrect source samples, error in data collection, or that genetic drift or selection played a role in shaping variation in the population. It is unlikely that dental morphological traits underwent selection in the African American population. The pattern of clustering in the PCA plot do not support errors in data collection, as such error should be random across samples. Instead, each sample is contained, and the plot shows three distinct clusters of African Americans, a distinct cluster of Africans, and a distinct cluster of Europeans. We hypothesize that a combination of admixture and drift shaped the pattern of diversity in African Americans.

The African sample clusters in the bottom right of the PCA plot, while the European sample clusters in the upper left. All African American samples fall along a diagonal between them. This diagonal, from top left to bottom right is the axis of admixture. The orthogonal axis, from bottom left to top right is the axis of a different evolutionary process, most likely drift.

The Gullah and UT samples nearly completely overlap each other on both the admixture and drift axes. Although these samples are from different time periods, there may have been gene flow between these geographic locations due
to proximity, which would explain the clustering of these two samples. NYU, USC, and UW samples overlap each other and form a separate cluster. These samples are all from the late 20th century, toward the end of The Great Migration in which African Americans moved left The South, and moved to cities in the northern and western parts of the country. This movement of people from a single region may explain the clustering of these three samples.

While the Freedman’s Cemetery was indistinguishable from the other samples in the STRUCTURE analysis, this sample formed a distinct group in the PCA. For example, the Freedman’s Cemetery sample falls in the same place along the axis of admixture as the UT sample, which is why these samples are indistinguishable in the STRUCTURE analysis. However, there is virtually no overlap between these samples along the orthogonal axis. This suggests that African Americans in Dallas experienced a different history than the African Americans in our dataset from other locations of the US despite the overlap in ancestry proportions. North Dallas Freedman’s town was established soon after the 14th Amendment was ratified, Freedman’s Cemetery served as the burial ground for the population of North Dallas Freedman’s Town between 1869 and 1907 (Davidson, 1999). Therefore, people buried there predated the Great Migration, which began in 1916. This time depth difference, along with the geographic distance from the other sampling locations, may explain why the Freedman’s Cemetery sample is separate from the contemporary samples used in our analyses along the axis of drift.
Conclusions

Our analyses confirmed that dental morphology can be used to evaluate mean differences in ancestry among samples. We found that reducing the amount of missing data in our sample increases accuracy in individual cluster estimates, and reduces the 90% credible regions. Our analyses demonstrated that ancestry estimation alone is not sufficient to understand the pattern of variation in a structured population. Using a combination of ancestry estimation and PCA, we found variation among our African American samples, and evidence that suggests African Americans in different locations and at different times have experienced distinct patterns of admixture and likely drift.
CHAPTER 5. SUMMARY AND CONCLUSIONS

My goal in this dissertation was to gain a better understanding of the history of African American admixture. Admixture has been ubiquitous in the history and prehistory of our species, and the dynamics of those past events have no doubt been affected by the cultural and social context in which they occurred. African Americans provide an example of this process that comes with a well-documented history of social and cultural pressures. These pressures have not only resulted in health and social disparities, but has also played an unconscious role in the study of this population. My point of entry into this field of study was the realization that previous studies of African American admixture were limited in important ways. As I mentioned in the Introduction, only two models of admixture had previously been tested on samples of African Americans; both were overly simplistic (Glass and Li, 1953; Long, 1991; Parra et al., 2001b; Pfaff et al., 2001; Gravel, 2012). Both models include Africans and Europeans coming together one time in the past to form the first generation of African Americans. In one model, all subsequent generations were the result of unions between African Americans. In the other model, Europeans continued to contribute to the African American population each generation after the initial union of Africans and Europeans, but Africans did not. These models ignore the continuous influx of Africans to the region that is now the US during and after slavery (Curtin, 1969; Gibson and Lennon, 1999). These models also assume
that random mating occurred in a single, unstructured African American population over the past 400 years.

In Chapter 2, I tested these admixture models, along with two models that account for the Africans who immigrated to the US over the course of African American history. The rate of African contribution in these models was based on the history of African immigration as reported in the US Census and importation records. These models fit two of the samples used in that analysis better than either of the simple models. As discussed in the chapter, the models involving continued African contribution did not fit best to the third sample due to 1) the sampling scheme, which did not allow for contribution from Africans or Europeans in the past two generations, 2) variation in admixture history among regions of the US, 3) lack of statistical power in the relatively small sample, or 4) some combination of these factors. Nonetheless, the simplest model of “one time” admixture, which is often assumed in studies of African Americans, was rejected for all samples.

The novel contributions of this chapter are two-fold. First, incorporating even a simplified history of African Americans in which Africans and Europeans continue to contribute over time improves the fit of the model for some samples. Second, the European contribution to the African American population was higher than that estimated under the two unrealistic admixture models. This underestimate of the European contribution is a direct consequence of disregarding continuous contribution from immigrating Africans. Including
continued contribution from Africans necessitates a higher European contribution to result in the same population average ancestry observed today.

Consistent with previous studies (Bryc et al., 2015; Baharian et al., 2016), my second chapter demonstrated variation in African American ancestry among samples from different US regions. Lacking in those previous studies was an attempt to understand the causes of this geographic variation in ancestry. Given the history of slave importation, anti-miscegenation and segregation laws, the large-scale movement of African Americans from the rural South to the industrial North and West during the Great Migration, and the Civil Rights Movement, I hypothesized that demographic, social, and political change would result in not just geographic but also temporal variation in European and African contributions to African Americans. The remainder of my dissertation focused on understanding the evolutionary forces that produced this geographic and temporal variation in the admixture process in African Americans.

Dental morphological data are situated geographically and temporally. The cemeteries from which the historic samples used in my analyses are derived are well documented. For example, for the Freedman’s Cemetery sample, the information available on the origin of the sample includes the dates of burial and the social dynamics between the people buried there and the surrounding area. African Americans in the Dallas area were relegated to North Dallas Freedman’s Town by the Fall of 1865 Vagrants Ordinance, which was instated to prevent African Americans from settling in Dallas proper. Freedman’s Cemetery served
as the burial ground for the population of North Dallas Freedman’s Town from 1869 to 1907 (Davidson, 2004, 2007; Turkel, 2005). These data provided a potential way to look back in time at the variation that existed in prior to the Great Migration and Civil Rights Movement. However, before I could begin analyzing and comparing ancestry in dental samples, I had to assess the utility of dental data for ancestry estimation. Therefore, my goal in Chapter 3 was to formally assess the informativeness of dental morphological data in ancestry estimation.

In this study, I showed that dental morphology data do in fact contain information about the admixture process. Evidence for this conclusion, quoting from the chapter, is that, “(a) African and European individuals had consistently higher membership in one cluster than the other, and (b) African American individuals had consistently higher membership in the same cluster in which African individuals had high membership.” Additionally, I concluded that dental morphological data, particularly in nondichotomized form, show a signal of ancestry and are useful in the investigation of inter-population differences.

In Chapter 4, I examined the relationships between geographic location, time, and mean biogeographic ancestry using nondichotomized dental morphological data in six samples of African Americans from across the US. My goal was to understand how differing histories within the US affected patterns of admixture in African Americans. Consistent with other studies of African American ancestry (Bryc et al., 2015; Baharian et al., 2016; Gross, 2018), the
combination of STRUCTRE and PCA analyses presented in this chapter showed geographic variation in African and European ancestry.

A novel finding is that there is also temporal variation in amounts and patterns of admixture. This conclusion is based on PCA analysis (see Figure 4.2), which was used to summarize trait frequency variation among the samples. The analysis showed a unique pattern of trait frequency variation in the Freedman’s Cemetery sample compared to the other African American samples. Trait frequencies in African Americans fall between those of Africans and Europeans, as expected following admixture. However, the African American samples are also distinct from one along the orthogonal axis, as might be expected if drift has shaped trait frequencies in the samples. The Freedman’s Cemetery sample is particularly distinct from the other samples on this orthogonal axis. This separation constitutes another novel finding that PCA can be applied to dental morphology to jointly examine drift and admixture.

Concluding statement

When I began this research, my desire was to address the issues inherent to previous studies of African American admixture. As stated above, only unrealistic models have been tested, and these models exclude any consideration of the African contribution to the African American population after the first generation. This omission may derive from the assumption that Africans
and African Americans are the same, and has led to the prioritization of Europeans in the study of African American admixture.

The racialized conflation of Africans and African Americans, though African Americans are a distinct population, perhaps derives from the “one-drop rule”, and therefore Africans become African Americans upon arrival to the US. The “one-drop rule” has its roots in the long history of slavery and Jim Crow segregation, and dictates that any person with any known African ancestry is considered black, and thus the same subjugated race (Myrdal, 1944; Berry and Tischler, 1978; Davis, 2001). As a result of considering African immigrants to be of the same group as African Americans (or vice versa – that the African ancestry of African Americans dictates their standing in the US), there is a Eurocentric view of African American admixture in which only Europeans contribute, and no acknowledgement of continued gene flow from Africans.

I have attempted to better understand admixture in African Americans, which is one of the myriad such instances that have occurred over the course of our species’ existence (Pickrell and Reich, 2014; Reich, 2018). Admixture throughout human history has molded diversity in our species, and it is the only evolutionary process that never occurs in a cultural vacuum. It is unclear what social and cultural conditions mediated most of these admixture events, but this is not true for African Americans. There is a plethora of information on African American history, e.g., slave importation records, various laws, and anecdotal accounts, that provide a cultural context for understanding this particular instance.
of admixture. However, we still have surprisingly little understanding of how social and cultural rules have structured the pattern of admixture. While I attempt to address the issues inherent to previous studies of African Americans, I realize that there is still much work to do to understand this population, and its history of admixture. Currently, there are contemporary genetic samples from African Americans across the US. However, there is a paucity of historic samples, whether dental or genetic, against which they can be compared. In the future, I hope to obtain such historic samples to continue to explore how diversity in African Americans has changed over time as a result of the social and cultural pressures they have experienced.
APPENDIX

Data descriptions

Table A.1. Dental morphology data description.

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample ID</th>
<th>n</th>
<th>Geographic Origin</th>
<th>Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>European</td>
<td>Europe</td>
<td>111</td>
<td>Berin, Cologne, Gottingen, Wurttemberg, Schleswig-Holstein, Osterberg, Prachim, Paris, Tronohje, Lund, Schwiz Canton, Trent, Ursern</td>
<td>American Museum of Natural History</td>
</tr>
<tr>
<td>US (European Americans)</td>
<td></td>
<td>224</td>
<td>Cleveland (Bolton-Brush), Rome, NY (Oneida Poorhouse), Albany</td>
<td>Case Western Reserve University, State Museum of New York</td>
</tr>
</tbody>
</table>

African American

<table>
<thead>
<tr>
<th>Population</th>
<th>n</th>
<th>Geographic Origin</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>NYU</td>
<td>67</td>
<td>New York</td>
<td>NYU Dental School</td>
</tr>
<tr>
<td>USC</td>
<td>65</td>
<td>Southern California</td>
<td>USC Dental School</td>
</tr>
<tr>
<td>UT</td>
<td>100</td>
<td>Tennessee</td>
<td>UT Health Sciences Center</td>
</tr>
<tr>
<td>UW</td>
<td>34</td>
<td>Washington</td>
<td>University of Washington</td>
</tr>
<tr>
<td>Freedman’s Cemetery</td>
<td>332</td>
<td>Dallas</td>
<td>Texas Department of Transportation</td>
</tr>
<tr>
<td>Gullah</td>
<td>55</td>
<td>South Carolina/Georgia</td>
<td>Ohio State University</td>
</tr>
</tbody>
</table>

Table A.2. Genetic data description.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Population</th>
<th>n</th>
<th>Geographic Origin</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNPs</td>
<td>African</td>
<td>112</td>
<td>Yoruba</td>
<td>HapMap</td>
</tr>
<tr>
<td></td>
<td>European</td>
<td>110</td>
<td>Utah</td>
<td></td>
</tr>
<tr>
<td></td>
<td>African American</td>
<td>49</td>
<td>SW USA</td>
<td></td>
</tr>
<tr>
<td>STRs</td>
<td>African</td>
<td>50</td>
<td>Yoruba</td>
<td>Tishkoff et al., 2009</td>
</tr>
<tr>
<td></td>
<td>European</td>
<td>29</td>
<td>France</td>
<td></td>
</tr>
<tr>
<td></td>
<td>African American</td>
<td>98</td>
<td>Baltimore, Chicago, North Carolina, Pittsburgh</td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


McLean DC, Spruill I, Argyropoulos G, Page GP, Shriver MD, Garvey WT. 2005. Mitochondrial DNA (mtDNA) haplotypes reveal maternal population


Voyages Database. 2009. Voyages: The Trans-Atlantic Slave Trade Database.
Available from: http://www.slavevoyages.org/assessment/estimates