EXPLORING THE RELATIONSHIP BETWEEN CRANIAL NON-METRIC AND

METRIC TRAITS FOR ANCESTRY ESTIMATION

by

Felicia S. Wun, B.S.

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Arts
with a Major in Anthropology
December 2014

Committee Members:

M. Kate Spradley, Chair

Elizabeth Erhart

Joseph Hefner

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Ancestry estimation is an important component in the discipline of forensic anthropology. Forensic anthropologists either visually assess skeletal remains through cranial macromorphoscopic traits or via craniometric analyses. Typically these two approaches are assessed separately as standalone methods. In this study, a variety of machine learning methods (decision tree analysis, random forest modeling, artificial neural networks, support vector machines, and linear discriminant functions) were applied to macromorphoscopic, craniometric, and combined (macro and metric) datasets to evaluate the classification accuracies of each and to explore how their individual and combined contributions may affect the estimation of ancestry.

Overall, the random forest model performed the best out of the methods in two of the datasets with a classification accuracy of 95% for the metric data and 90% for the macromorphoscopic data. For the combined dataset, the support vector machines performed the best at 90%. The present study demonstrates the utility of these new methods contributing a greater wealth of information to group classification and also improving knowledge that these two data types can be combined into a single statistical analysis with classification accuracies of 90% and above for specific machine learning methods.

I. INTRODUCTION AND LITERATURE REVIEW

Within the field of forensic anthropology, the biological profile is created to provide estimations of age, ancestry, sex, and stature for unidentified human skeletal remains (Christensen et al. 2014; Gill 1998; Ousley et al. 2009). Estimating ancestry is a key component to informing additional aspects of the biological profile by relating a group of biological traits to a socially constructed race category as a means of providing further evidence in the identification of an individual (Anderson 1998; DiGangi and Hefner 2013; Klepinger 2006; Sauer 1992; SWGANTH 2013). Methodologies using visual assessments and established measurements of cranial and postcranial elements provide insight into the degree of varied phenotypic trait expression among diverse population groups (Brues 1990; DiGangi and Hefner 2013; Hefner et al. 2012; Gilbert and Gill 1990).

Forensic anthropologists traditionally utilize non-metric traits (qualitative) through visual assessments or standardized measurements and statistical analyses (quantitative) to guide their estimation of ancestry (Birkby et al. 2008; Hinkes 1993; Hefner 2009; Hurst 2012; Jantz and Ousley 2005; Rhine 1993; Rhine 1990; Ousley et al. 2009; Sauer and Wankmiller 2009; Spradley et al. 2008; Wescott 2005). Information gleaned from both non-metric and metric assessments contributes a wider range of understanding when assessing skeletal remains of an unknown individual (Birkby et al. 2008; Hefner et al. 2014). Oftentimes, visual assessments for estimating ancestry are problematic because of their typological, classificatory nature. The most commonly used practice in the assessment of ancestry using non-metric traits in the United States

typically requires individuals to be classified into three discrete groups: White, Black, or Native American/Hispanic/Asian.   In most instances, the visual assessment of ancestry places an unknown individual within a neat box while disregarding the nuances of variation. Metric analyses utilize standard cranial measurements and are used within a statistical framework to estimate ancestry providing an estimate expressed as a posterior probability (Christensen et al. 2014; DiGangi and Hefner 2013; Hefner 2007).   Because of their historical roots in typology, forensic anthropologists must be careful with their visual assessments of ancestry.

This current research seeks to provide a better understanding of the two methods used to estimate ancestry and quantify their individual and collaborative contributions in their applications in medico-legal contexts. This research aims to present a historical and temporal survey of both of these methods, an undertaking that has not been fully explored within the preexisting literature to fill in the gaps charting the development of the two common methodologies for ancestry estimation: non-metric and metric. Although both methodologies are available for use by a forensic anthropologist, the craniometric method is further developed than the macromorphoscopic method through the inclusion of multiple reference samples attributed to the Forensic Anthropology Data Bank (FDB) (Jantz and Moore-Jansen 1984). This discussion will provide new insights into the current state of ancestry estimation and develop an understanding surrounding previously established methodologies. The outcome of this research not only informs how forensic anthropologists currently apply and understand the contributions of these two methods based on cranial morphology and overall size but also expands upon the preexisting literature that is currently available. This will guide current practitioners to understand

how and why the contribution of both non-metric and metric methods impacts ancestry estimation. Lastly, the research will be significant because all samples are derived from a matched data set. Otherwise known as a paired data set, the non-metric and metric data have a one-to-one relationship meaning that each data source contains the same individuals in the sample who have corresponding non-metric and metric information between the two data sets (D'Orazio 2014). The advantage to having statistical matching is the ability to investigate the relationship of the two variables within each data source at a micro or macro level. Therefore, each individual that is compared in this research has both corresponding non-metric and metric data whose relationship will be analyzed.

Application of Non-Metric Analyses for Ancestry Estimation in Forensic Anthropology

Forensic anthropologists utilize skeletal morphology as part of the biological profile to provide further information correlating a predicted social race from visual assessments and/or measurements. (Klepinger 2006; Sauer 1992). These broad groupings may be useful in the forensic context; however, the true nature of biological variation may be concealed (Relethford 2009). Certain non-metric traits in the mid-face have been isolated by practitioners based on their perceived contribution to the overall estimation of ancestry (Brues 1990; Hefner et al. 2014). To effectively apply these traits to ancestry estimation, the "trait list method" was introduced to provide a comprehensive collection of individualized traits that corresponded to, and were representative of, distinct population groups (Hughes et al. 2011; DiGangi and Hefner 2013). A particular trait expression would more likely be expected in one population than the other based on calculated trait frequencies of each individual trait (Gill 1998; Hughes et al. 2011; Rhine

1990).   For example, in White, Black, and Native American groups, a suite of traits specific to each group is provided. However, there is still a great deal of overlap in non-metric trait expression among the groups and one individual does not fit neatly into one group (Hefner 2009; Rhine 1990).

Earnest A. Hooton, a professor of Anthropology at Harvard University, created an early, standardized form of the trait list called the "Harvard List," which was used as an exemplar for practicing forensic anthropologists for many years. Hooton established the "Harvard List" to synthesize the gamut of non-metric and metric traits that were especially helpful during skeletal analysis (Ferguson et al. 2011; Hefner et al. 2012). As a professor at Harvard University, Hooton collected both non-metric and metric data from a variety of sources, including his work at the Pecos Pueblo archaeological site and from criminals in the US (Hefner 2007; Rafter 2004). Hooton preferred using non-metric traits due to their relative ease of use and the ability to evaluate crania based on the qualitative nature of the trait manifestations (Hefner et al. 2012; Hooton 1926). Each trait was scored based on the expression of various character states for each trait or based on its presence or absence (Hefner et al. 2012). The entirety of the analysis was based on the overall experience of the observer in observing the morphologies of the crania and the qualitative nature of the descriptors used (Hefner et al. 2102; Hooton 1926; Wheat 2009). The apparent subjectivity in this type of non-metric assessment led Hooton to emphasize the need for standardization of non-metric traits. Many times, the *Gestalt* technique was employed by physical anthropologists. The *Gestalt* refers to the ability of an experienced practitioner who is conditioned by years of training and experience to quickly formulate an assessment of sex or ancestry for an unidentified set of remains (Hefner 2009; Hefner

and Ousley 2006). In ancestry estimation, the *Gestalt* is noted when a forensic anthropologist immediately classifies a set of unidentified remains through a basic visual assessment but can not verbalize exactly what led them to their conclusions. Therefore, factors such as education, experience and training must play a role in influencing and shaping an individualized assessment of the cranial *Gestalt* (Wheat 2009).

Furthermore, the expansion of later research to understand just how well these non-metric traits performed for estimating ancestry generally focused on detailing non-metric traits that were representative of the different population groups at hand with their accompanying trait frequencies. However, misclassification rates were not assessed and there was a lack of concrete research evaluating these traits in a more robust, statistical framework to echo Hooton's need for proper standardization. Instead research involving the publication of trait lists including those from Rhine (1990) and Gill and Rhine (1990) were disseminated and widely used for consultation during forensic casework (Christensen et al. 2014; Hefner et al. 2012; Hughes et al. 2011).

Currently, to visually assess ancestry, the observer evaluates the skull noting various features. Typically a reference guide displaying the non-metric traits for each population group is consulted during the analysis. The observer then attempts to associate the skull to a particular population group based on the recorded traits. At times, the feature would fall into the intermediate category of trait expression leaving the forensic anthropologist with the task of choosing between two population groups. Although standardized trait lists are still consulted, there is no universal manner of collecting non-metric information for ancestry in forensic anthropology. However, several approaches towards a systematic form of non-metric trait collection have been

developed and thus, are beginning to be implemented into practice (Hefner 2009; Osteoware 2012).

The future of qualitative analysis for the estimation of ancestry has been reinvigorated by research involving macromorphoscopic traits of the cranium in forensic anthropology (Digangi and Hefner 2013; Hefner 2009; Hefner and Ousley 2014; L'Abbé et al. 2009; Ousley and Hefner 2005). These traits are "quasi-continuous" traits of the cranium that reflect the soft tissue variations in the skin (Ousley and Hefner 2005; Hefner et al. 2012:295). For example, anterior nasal spine is not characterized by presence or absence but on an ordinal scale with varying degrees of trait expression based on three character states (1: Slight; 2: Intermediate; 3: Marked). Individuals from various populations are scored on this continuum of trait expression instead of assigning a dichotomous score of present versus absent for a particular trait. A trait may be present; however, the degree of trait expression will vary from individual to individual contrary to previous approaches of relegating a suite of traits whose extreme manifestations are immediately associated with a particular group. Hefner examined the correlation between macromorphoscopic traits that captured a wide pattern of variation amongst the samples (2009). Coupled with statistical analysis and examination of trait frequency distributions, these traits prove useful to studies examining biological variation amongst populations from this approach (Hefner 2009; Hefner et al. 2014; Hefner and Ousley 2014).

The inclusion of the Macromorphoscopic module in the Osteoware data collection program systematically incorporates widely used traits found in the Hefner (2009) study with the addition of other macromorphoscopic traits. The Osteoware data collection

program was created in partnership with the Smithsonian Institute as a means for standardization of data collection procedures (Osteoware 2011). Free and available to interested researchers, this software program incorporates standardized traits and measurements found in Standards (Buikstra and Ubelaker 1990) and acts as a standardized data entry tool for the collection of qualitative and quantitative observations of skeletal remains (Osteoware 2011). The Macromorphoscopic module includes both comprehensive descriptions and illustrations of each trait to guide observers during their analyses (Hefner 2012). An additional study by Klales and Kenyhercz investigated the reliability of ancestry estimation using these macromorphoscopic traits to determine the trait frequencies within their sample and the degree of their variation and also assessed interobserver error using the Macromorphoscopic module of Osteoware (Hefner 2012). Their results reinforced the notion that certain macromorphoscopic traits do not characterize certain population groups and that macromorphoscopic traits can be analyzed statistically to estimate ancestry - a method that is both scientific and repeatable (Klales and Kenyhercz 2014). Current research contributed to rectifying issues regarding subjectivity and lack of standardization of non-metric techniques to assure that these methodologies be constantly reevaluated (Christensen and Crowder 2009; DiGangi and Hefner 2013; Hefner et al. 2012; Vitek 2012). Future studies utilizing these types of approaches should be conducted and expanded to all available skeletal populations.

Application of Craniometric Analyses for Ancestry Estimation in Forensic Anthropology

Craniometric traits used for ancestry estimation rely on standardized measurements of the skull and are coupled with detailed definitions, precise

instrumentation, and multivariate statistical analysis (Hefner et al. 2012). Much of the

research related to ancestry estimation has been geared more towards craniometric data

because of its methodical nature and ability to perform under statistical analyses. Several

key studies and resources have shaped the nature of research for craniometric analysis in

ancestry estimation throughout the course of forensic anthropology. A study by Giles

and Elliot (1962) utilizing nine craniometric traits to evaluate ancestry using discriminant

function analysis was performed to classify crania based on the three group model

(White, Black, and Native American). Additionally Giles and Elliot provided

classification rates to inform just how well the discriminant functions performed in their

assessment.

Reference group samples used in craniometric analyses generally have much

more variety than their non-metric counterparts (Jantz and Moore-Jansen 1984; Howells

1973; Howells 1989. Additional forensic craniometric and demographic data of

identified or soon to be identified individuals were curated and housed in a database

entitled the Forensic Anthropology Data Bank (FDB) providing reference material for a

variety of individuals from various backgrounds (Jantz and Moore-Jansen 1984). The

introduction of FORDISC (Jantz and Ousley 2005) a discriminant function computer

program created by and Richard Jantz and Stephen Ousley elevated the previous methods

that utilized craniometric analysis by classifying an unknown individual based on the

similarity of their overall measurements compared with an expanding reference database

sample of known individuals (Christensen et al. 2014; Ousley and Jantz 2005). Including

the FDB, FORDISC also utilizes the craniometric data from William W. Howells that

includes measurements from 28 populations (Christensen et al. 2014; Howells 1973;

Howells 1989). Therefore, craniometric analyses have more reference groups for

assessing the ancestry of an unknown individual compared to the traditional three-group

model used in qualitative analysis.

Impactful strides related to craniometric analyses and ancestry estimation has

been made throughout the course of forensic anthropology's history. The addition of

reference groups into the FDB create a better, more diverse picture of the various,

existing population groups that are used for comparison when trying to attain the identity

of an unknown individual.

Exploring the Interaction of Cranial Non-Metric and Metric Traits in the Literature

Both cranial non-metric and metric data individually contribute information with

regards to ancestry estimation, however the relationship between these two data types has

not been extensively explored within the literature. Much of the early literature exploring

the relationship of cranial non-metric traits and metric variables utilized both primate and

human crania (Cheverud et al. 1979; Carpenter 1976; Corruccini 1974; Corruccini 1976;

Richtmeier et al. 1984; Wilson 2010) to explore the sources of non-metric trait

manifestation and metric variables and their relationship with one another. Although

these studies did not directly address issues in forensic anthropology, their findings

contributed to the development of ancestry estimation. The majority of the authors

focused on the epigenetic basis for the etiology of the non-metric traits by examining

their concordance with facial dimensions. Richtsmeier et al (1984) examined the effects

of both metric and non-metric traits in a sample of rhesus macaques. Fifty-six

craniometric variables and nine non-metric traits were used (Richtsmeier et al. 1984).

Specific non-metric traits were chosen to show the corresponding relationship with non-metric traits and their localized measurements.

In an additional study, the findings suggested a positive correlation between cranial non-metric traits and metric variables. Specifically, there were individual correlations between specific non-metric traits and metric variables, including minimum frontal breadth and metopism and nasal breadth with nasal spine sharpness (Carpenter 1976:340). Corruccini (1976) found a significant correlation between non-metric traits and metric variables; however, the author did not specifically look at non-metric traits and their corresponding metric measurements that contributed to overall cranio-facial variation. He proposed another explanation that variation in both non-metric traits and metric variables could be dependent upon another factor, not yet explored (Corruccini 1976).

Cranial non-metric traits and metric variables originate from a common point of development (Cheverud et al. 1979). The overall size and shape of crania is correlated with the presence and absence of certain non-metric traits (Cheverud et al. 1979). Wilson (2010) examined the correlation between cranial non-metric traits and metric variables using a sample of 20 crania from South East Asia and did not find a correlation between these two. However, she found 14 inter-trait correlations and three associations between single measurements of metric variables and non-metric traits (Wilson 2010).

The major questions in this research were two-fold: if the non-metric traits of the skull were actually an expression of size and if they were correlated to other non-metric traits that were assessed (Wilson 2010). Multiple traits that are expressed from the same gene will not be genetically distinct. Every metric and non-metric trait utilized was taken

from Buikstra and Ubelaker (1994). Results showed no correlation between non-metric

traits and metric calculations for each part of the skull. There were significant

correlations associated with certain non-metric traits and single cranial measurements.

Fourteen inter-trait correlations and three correlations between single measurements and

non-metric measurements were found (Wilson 2010). Further research within the same

population was recommended.

In a recent study, Hefner and colleagues (2014) examined the relationship

between cranial macromorphoscopic and craniometric data in White, Black, and Hispanic

population groups using random forest modeling. The Hefner et al. (2014) study is

methodological, emphasizing the use of random forest modeling to evaluate qualitative

and quantitative data to avoid violation of certain statistical assumptions that one may

encounter using linear discriminant function analysis for ancestry estimation.   The goal

was to provide an alternative method combining metric variables and nonmetric traits

into a single, stable model for analysis in lieu of the more traditional linear discriminant

function analysis. Random forest modeling successfully integrated both trait types within

a single analysis and opened up a discussion for future research regarding the utility and

comparability of using both macromorphoscopic and craniometric data for the assessment

of ancestry bringing together these two trait types within a single analysis. Although

research has been conducted by evaluating the contributions of both non-metric and

metric analyses, the contributions of both methodologies should be reevaluated to gauge

their relevance and helpfulness for ancestry estimation in forensic contexts. Specifically,

current methods using the macromorphoscopic traits outlined in the Osteoware program

(2011) and standard craniometric analysis were evaluated in their study.

For the purposes of this research, the term macromorphoscopic will be used throughout the remainder of the text, specifically in the results and discussion section, to discuss the traits used in the present research. The term macromorphoscopic was selected because it encompassed the "quasi-continuous" nature of the traits that were used in this study (Ousley and Hefner 2005; Hefner et al. 2012). Because the non-metric variables are selected from Hefner (2009) and the Macromorphoscopic module of Osteoware (Hefner 2012; Osteoware 2011), it is important to maintain parallel vocabulary to avoid confusion between the various terminologies used for cranial non-metric traits or discrete traits. To capture the entire picture of the continuum of trait expression amongst the sample populations, macromorphoscopic traits that could be expressed as various character states illustrating the range of variation were selected as well as those that were binary such as nasal overgrowth. Therefore, qualitative traits such as the macromophoscopic traits are the focus due to their contribution to ancestry estimation and forensic anthropology.

The field of forensic anthropology is constantly expanding and traditional methods such as the visual assessment of non-metric traits require strong statistical backing to address the standards set by the Daubert ruling (Daubert 1993). Forensic anthropologists look to other disciplines for innovative ways of data analysis that may strengthen current methods. Dating mining techniques uncover and interpret patterns within the data provided by building models to make predictions (Williams 2011). Therefore, the models that are created become another way that the data's meaning is physically expressed. These techniques are applied to a variety of disciplines including business, politics, medicine, engineering, and biology, to name a few (Williams 2011). It

is not surprising that these innovative techniques have been applied to forensic anthropology.

Within the preexisting literature there are a handful of studies that incorporate machine-learning techniques to answer questions relating to forensic anthropology (Corsini et al. 2005; Du Jardin et al. 2014; Hefner and Ousley 2014; Hefner et al. 2014; Mahfouz et al. 2007; McBride et al. 2001; Navega et al. 2014; Navega et al. 2013). Navega and colleagues (2014) used a variety of machine-learning methods (decision trees, artificial neural networks, statistical learning, instance based learning, and probabilistic learning) using tarsal bones for sex estimation. In this study, the "tree based algorithms and variants of the naïve Bayes algorithm" were the best at estimating sex with classification accuracies of 0.883. An additional study by Navega et al. (2013) introduced a program called AncesTrees to evaluate metric variables for ancestry estimation. Navega and colleagues applied random forest algorithms to create multiple randomized decision trees to accurately assign class membership for ancestry estimation.

The purpose of this study was to bring together non-metric and metric data within a single analysis using various machine-learning techniques (decision trees, random forest modeling, artificial neural network, support vector machines, discriminant function analysis) to estimate ancestry. Unlike previous studies, this study looks at the classification accuracies for the estimation of ancestry and how well these machine-learning methods perform with individual datasets (non-metric and metric data) and with a combined data set. The utility of combining traditionally separate statistical analyses using both non-metric and metric data in a single analysis was assessed. The combined analysis contributes a new way of evaluating ancestry.

Furthermore, the present research contributes to the field of forensic anthropology in three ways. First, this research contributes to the preexisting literature that explored the utility of both macromorphoscopic and metric data within the context of biological and forensic anthropology. The present study builds upon and further dissects the relationship between macromorphoscopic and metric traits and their relative interaction to ancestry estimation and with one another. The Wilson (2010) and Hefner et al. (2014) studies establish a foundation for new inquiry surrounding the relationship between the two trait types and the associated methodologies for combining them within a single analysis, respectively. However, within the context of forensic anthropology it will be important to investigate how the methodologies from these two trait types perform in collaboration or provide a better estimation of ancestry individually rather than combined in a single analysis. Second, the current research also provides new data, building upon data for trait frequencies of Hispanic crania leading to a contribution to the preexisting collection of macromorphoscopic data in order to identify patterns of variation between individuals. Skeletal collections in the United States are predominantly composed of White and Black individuals (Hunt and Albanese 2005; Hunt and Spatola 2008). Therefore, much of the preexisting literature is limited due to the availability of sample populations used in research. The addition of macromorphoscopic data on Hispanic populations will be key. Third, the inclusion of macromorphoscopic data of a Hispanic cranial sample is significant due to the recent influx of Mexico-United States border related deaths and the investigations geared towards that cause (Anderson 2008; Spradley 2008; Spradley 2013). The inclusion of population specific data will help to inform other aspects of the biological profile and existing methodologies associated

with the biological profile to expand and inform current best practices. By utilizing a

broad-based approach to investigating the relationship between these trait types, the

current research not only focuses on the broader association between macromorphoscopic

and metric traits but looks at how the two trait types in conjunction, not individually

inform current practices in forensic anthropology and understanding variation.

II. MATERIALS AND METHODS

Materials

Cranial non-metric and metric data was collected from modern skeletal populations from the following research collections throughout July 2013: the Maxwell Museum at the University of New Mexico, the Pima County Office of the Medical Examiner (PCOME) in Tucson, Arizona, the William M. Bass Donated Skeletal Collection at the University of Tennessee-Knoxville, and the Texas State University Donated Skeletal Collection. Crania from three population groups were sampled for this study: American White, American Black, and Hispanic. Macromorphoscopic data from a total of 360 crania was evaluated by the author. The craniometric data was obtained from previously collected records through correspondence with the respective institutions. To complete a matched data set of the same individuals for both macromorphoscopic and craniometric analysis, 180 crania were ultimately used for this research.

Maxwell Museum Documented Skeletal Collection -University of New Mexico

As of 2013, the Maxwell Museum's Documented Skeletal Collection consisted of 278 indviduals of varying sex, age, and ancestry (UNM 2013). The sample was mainly comprised of self-donated or next of kin donated individuals (Komar and Grivas 2008; UNM 2013). Alternatively, unclaimed individuals have been obtained through the Office of the Medical Investigator or the Department of Anatomy at the University of New Mexico with known documentation (Komar and Grivas 2008). Most residents were

residents of New Mexico and ranged from fetal remains to decedents older than 65 years of age (Komar and Grivas 2008; UNM 2013). Sex and ancestry were self-reported in self-donation cases and reported by the family and/or death certificate in known cases (Komar and Grivas 2008). For unknown individuals, information regarding age, sex, and ancestry were gathered from the autopsy report and only age and sex were assessed for anatomical donations.

Macromorphoscopic data was collected from American White, American Black, and Hispanic crania. A total of 124 crania were examined. Two male crania had no associated population group, and were deemed as having "no data." One crania was excluded due to the highly fragmented nature of the remains and could not be visually scored. Because a matched dataset was used, only four crania were used in the analysis. (see Table 1).

Table 1. Crania sampled from the Maxwell Museum-University of New Mexico

	American Black	American White	Hispanic	Total
Male	0	0	0	0
Female	0	0	4	4
Total	0	0	4	4

Pima County Office of the Medical Examiner (PCOME) - Tucson, Arizona

From 2001 to 2007, over 1000 individuals died trying to cross the United States-Mexico border (Anderson and Parks 2008). The remains associated with these deaths were brought to the PCOME for post-mortem analysis. Individuals that cross the United States-Mexico border without proper documentation and of non-United States nationality

are designated "undocumented border crossers" or "UBC's" (Anderson and Parks 2008; Tise et al. 2013) by PCOME. A high percentage of the individuals associated with border related deaths have been identified as Mexican nationals (Anderson 2008; Tise et al. 2013). Due to the unknown nature of the remains, the sex, age, and ancestry of these individuals vary. Multiple lines of evidence are used to identify these undocumented border crossers including cultural artifacts that provide contextual clues to the identity of these remains including personal effects such as foreign documentation, foreign currency, and articles of clothing and/or mementos (Birkby et al. 2008). Thirty-six crania were evaluated based on availability of the remains at PCOME, with 35 being included as part of the matched dataset (see Table 2).

Table 2. Crania sampled from the Pima County Office of the Medical Examiner's Office (PCOME)

	American Black	American White	Hispanic	Total
Male	0	0	32	32
Female	0	0	2	2
Indeterminate	0	0	1	1
Total	0	0	35	35

<u>William M. Bass Donated Skeletal Collection - University of Tennessee-Knoxville</u>

As of 2011, the William M. Bass Donated Skeletal Collection consisted of a little under 1000 individuals of known sex, age, and ancestry. This collection consists of mostly American White males, however, American Black, Hispanic, and Asian individuals were also represented within the collection (Shirley et al. 2011). The majority of these individuals were residents of Tennesseee and reflected a varying age range from fetal to the majority of the individuals being between 50-101 years of age (Shirley et al.

2011). Therefore, the individuals in this collection tended to be associated with a population group of American White.

American White, American Black, and Hispanic crania were evaluated. A total of 139 crania were evaluated. To satisfy the conditions of the matched dataset, only 129 crania were included in the analysis (see Table 3).   Additionally there were 11 newer donations whose crania did not have an associated sex and ancestry. The macromorphoscopic observations for those individuals were omitted from the sample.

Table 3. Crania sampled from the William M. Bass Donated Skeletal Collection

	Black	White/EA	Hispanic	Total
Male	32	54	6	92
Female	4	32	1	37
Total	36	86	7	129

Texas State Donated Skeletal Collection - Texas State University

The Texas State Donated Skeletal Collection at the Grady Early Forensic Anthropology Research Laboratory (GEFARL) consisted of approximately 65 available skeletons and was comprised of self-donated individuals or next of kin donations of varying sex, age, and ancestry with the majority of the collection represesenting an older population of Texas residents (FACTS 2014; Mavroudas et al. 2012). Additionally, the collection contained a few cremated donations that were ommited from the sample. Crania of all available American White, American Black and Hispanic groups were visually assessed. A total of 41 crania were examined with 12 being included as part of the matched dataset. (see Table 4) Most crania were identified as American White due to the demographic background of the individuals that participated in the willed body

donation program.

Table 4. Crania sampled from the Texas State Donated Skeletal Collection

	Black	White	Hispanic	Total
Male	0	7	1	8
Female	0	4	0	4
Total	0	11	1	12

Cranial Non-Metric (Macromorphoscopic) Traits

A total of 11 non-metric, macromorphoscopic traits were evaluated for this study. The first ten macromorphoscopic traits were selected from Osteoware 2.4.037 within the Macromorphoscopic 1.28 module (Hefner 2012; Hefner 2009; Osteoware 2011). Emphasis was placed on traits that were located in the mid-facial region as that area has been consistently used in the evaluation of ancestry (Brues 1990; Gill 1998; Rhine 1990; Ousley and Hefner 2005) and palate shape was included as the final trait. The 11 macromorphoscopic traits and scoring ranges are listed below (see Table 5). These character states and trait descriptions originate from the definitions in Hefner (2009) and are also taken from the Macromorphoscopic chapter of the Osteoware Software Manual (Hefner 2012). Complete, detailed explanations of these traits and their corresponding degrees of character state manifestation are provided (see Appendix A).

Table 5. Eleven macromorphoscopic traits that were evaluated through visual assessment

Trait Cat Key	Trait	Scoring Range
ANS	Anterior Nasal Spine	1-3
INA	Inferior Nasal Aperture	1-5
IOB	Inteorbital Breadth	1-3
MT	Malar Tubercle	0-3
NAW	Nasal Aperture Width	1-3
NBC	Nasal Bone Contour	0-4
NBS	Nasal Bone Shape	1-4
Novg	Nasal Overgrowth	0-1
OS	Orbital Shape	1-3
TPS	Transverse Palatine Suture	1-4
PS	Palate Shape	1-3

Palate shape was included as the eleventh trait to be evaluated because of the absence of recent research in associating morphological attributes of the palate with ancestry. Instead, the focus of research has been directed towards using palate shape for sex estimation in recent years (Williams and Rogers 2006; Galdames et al. 2008). More importantly, palate shape is included on most non-metric trait lists for ancestry estimation associating particular shapes to correspond with particular population groups (Rhine 1990; Krogman and Işcan 1986). In most instances, palate shape is described as parabolic, hyperbolic, elliptical, small U or big U (Hooton cranial observations and indices; Hefner et al. 2012). In light of these subjective terms for describing palate shape,

there are difficulties for standardizing that information as most observations are based on the judgement of the observer. In response to this predicament, recent studies have addressed this issue by applying methods of digitization and metric analysis to assess this trait within a statistical framework (Burris and Harris 1998; Byers et al. 1997; Maier 2013).

To supplement the other macromorphoscopic traits in Osteoware (2011), palate shape and each of its three character states were formulated with detailed definitions (see APPENDIX B-C). The trait definition and its three character states were developed in collaboration with Dr. Joseph Hefner and Dr. Laurel Freas at the Joint POW/MIA Accounting Command's Central Identification Laboratory (JPAC-CIL). To collect palate shape, three character states are evaluated and scored from 1-3: elliptic (1), parabolic (2a and 2b), and hyperbolic (3). For the purposes of statistical analysis, the parabolic 2a and 2b scoring was collapsed to a general score of 2.

Although not integrated into Osteoware (2011), in time, this trait will eventually make its way into the Macromorphoscopic module for the continued collection of palate morphology. Palate shape is useful as a macromorphoscopic trait to be incorporated into Osteoware. Additionally, the collection of macromorphoscopic data is necessary for populations that are underrepresented within the preexisting literature. The difficulties of identifying specific morphologies representative of any Hispanic population will be further aided by the data collected from the Hispanic individuals from the Maxwell Museum, PCOME, University of Tennessee, and Texas State University.

Craniometric Data

A complete list of the measurements originating from Howells (1973) and additional craniometric measurements found in FORDISC (Jantz and Ousley 2005) with accompanying definitions is provided when available (see APPENDIX D). The author did not perform any craniometric measurements during the data collection process, therefore, the craniometric data originated from previously recorded measurements on file, either through instrument determination and/or digitization and assess was granted through correspondence with the respective institutions. The craniometric data for each individual were compared with the scores for the macromorphoscopic traits as a matched data set to evaluate the relationship between the macromorphoscopic traits and related areas for craniometric data to assess their respective contributions to the overall estimation of ancestry.

Methodology

Prior to data collection, the author spent time at the Joint POW/MIA Accounting Command's Central Identification Laboratory (JPAC-CIL) reviewing the Macromorphoscopic module of the Osteoware (2011) program with Dr. Joseph Hefner on sample crania that were part of a skeletal collection housed in the facility. Approximately 30 crania were evaluated during this time. None of these individuals were included as part of the overall sample. The purpose was for the author to review the traits and practice upon available crania. By doing so, the author had a period of time set aside to review and understand the different character states and visual manifestations of each trait. All 16 macromorphoscopic traits within the Osteoware module were evaluated

23

during this time. During the actual data collection period, the focus shifted to the ten

selected traits with the inclusion of palate shape.

A catkey was assigned to each institution for easy organization of data. A catkey

or catalogue number is a unique designation or label that corresponds to a selected

grouping within the sample. Catalogue numbers may consist of either letters or numbers.

For this research, the catkeys reflect the name of the particular skeletal collection from

which the sample crania were derived from. The catkeys established for samples from

each institution include

* UNM – Maxwell Museum Donated Skeletal Collection - University of New Mexico

* PCOME- Pima County Office of the Medical Examiner - Tucson, Arizona

* UTK – William M. Bass Donated Skeletal Collection - University of Tennessee-

Knoxville

* TSU- Texas State Donated Skeletal Collection - Texas State University

For example, an individual cranium from the Texas State Donated Skeletal Collection

would be labled with a catkey and donated number: TSU D15-2010. A catkey and

documented number were recorded along with the scores from Osteoware. Additionally,

any notes were taken down to explain any types of anomalies and/or pathologies, traits

that were unable to be scored, and anything else that was significant.

For each crania within the sample, each of the 11 macromorphoscopic traits were

evaluated and scored according to the interval provided in Osteoware (2011). Palate

shape was evaluated and recorded separately because it had not been incorporated within

the macromorphoscopic trait module in Osteoware (2011) during time of data collection.

For each trait, a score was selected in the Macromorphoscopic module of Osteoware (see

Figure 1).
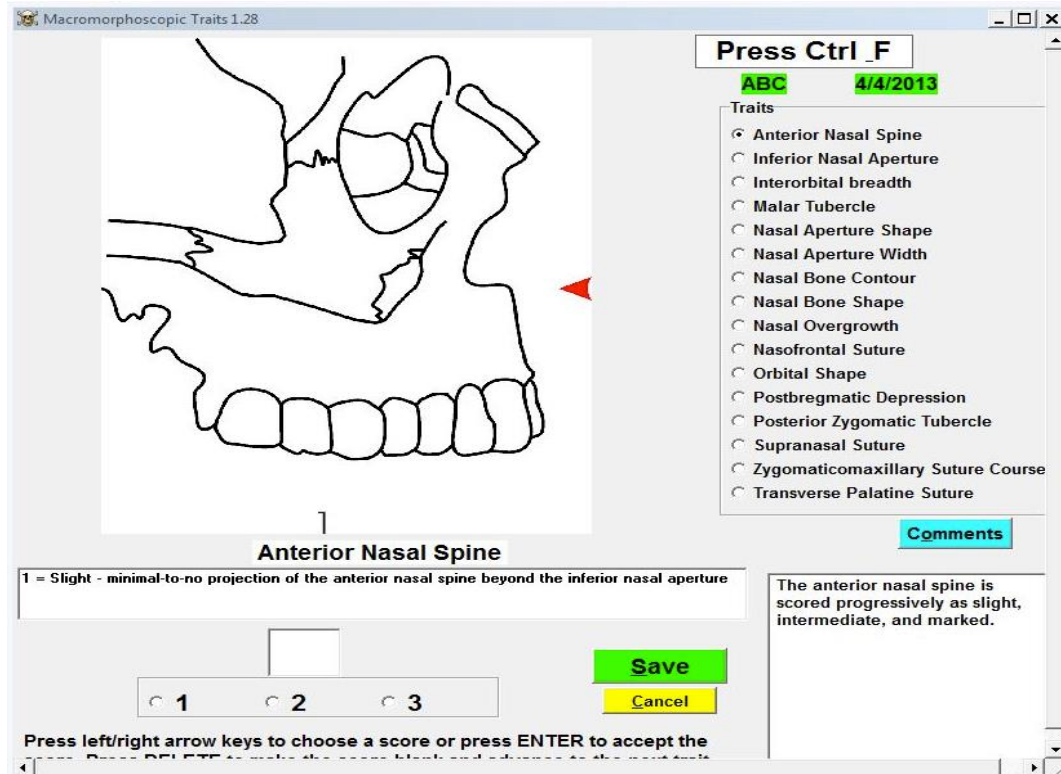
Figure 1. Screen capture of the Macromorphoscopic module in Osteoware 2.4.037 (2011)

Each score was manually recorded in a notebook and an additional spreadsheet

was created as an additional resource as a record of data collection for each individual.

For each trait, the definition was read twice and then the trait was examined for each

individual while consulting the illustrations in the Osteoware (2011) module during the

scoring process. After both definitions and illustrations were consulted, a score was

assigned for the trait for the ten traits in Osteoware (2011) for each individual. After the

scoring process commenced for each individual, the scores were saved into an Advantage

Data Architect database where the identification number and catkey for each crania was

entered. This process was repeated for every individual that was sampled. For traits that

were located in areas of severe damage, the score was left blank. Traits that garnered a

score of zero, based on the trait definition were manually added.

Due to time restraint, each sample was evaluated once. Multiple trials would

have been helpful to assess intraobserver error, however given the time restraints, this

could not be accomplished. Interobserver error was accomplished through the

comparison of the same individuals from the William M. Bass Donated Skeletal

Collection that had previously been examined during Hefner's study (2009) with the

present one. The author obtained a copy of the macromorphoscopic data for each

individual scored by Hefner from the Bass Collection and both scores from the author

and Hefner were subsequently used to assess interobserver error.

The macromorphoscopic samples were pooled by population group and given the

following catkey designations:

* WH – American White Individiual

* BLK – American Black Individual

* HSPN – Hispanic Individual

Statistical Analysis of Macromorphoscopic Data

Frequency distributions of all 11 macromorphoscopic traits were generated using the SPSS Statistics Package 22.0.0 for the three established population groups: American White, American Black, and Hispanic. These frequencies were compared to the trait frequencies found in Hefner (2009), Hurst (2012), and Rhine (1990) for further evaluation to explore possible differences between periods of data collection and from individuals derived from different population groups.

Interobserver error was conducted solely on the University of Tennessee sample with macromorphoscopic data previously collected and scored by Dr. Joseph Hefner. Subsequent scoring from the author and Hefner was compared to evaluate what traits were easiest to score and replicate and which traits were difficult to evaluate. An inter-rater reliability analysis using the Fleiss' kappa statistic was performed to determine consistency among raters. The Fleiss' kappa statistic is used for any number of observers using categorical data for a specific fixed number of items (Fleiss 1971).

Application of Data Mining Techniques

Data mining techniques can be applied to gain insight and knowledge from data. These techniques are expressed through model building used to interpret patterns and relationships in the data (Williams 2011). Data mining techniques may be used for classification problems or regression, but usually they consist of three steps: 1) "initial exploration;" 2) "model building or pattern identification with validation/verification;" and 3) "deployment" (StatSoft 2014). The first step includes data preparation such as data cleaning and any exploratory research. Second, the appropriate model for assessing

the data is selected with a particular eye to which model has the best predictive power.

Third, the chosen model is applied to new data to assess how well it predicts outcomes.

In this study, dating mining techniques for classification were applied to the

macromorphoscopic and craniometric data. These analyses included decision tree

analysis, random forest modeling (RFM), artificial neural networks (aNNs), support

vector machines (SVM), and linear discriminant function analysis (LDF). Each

technique was applied three times: once to each of the macromorphoscopic and

craniometric data separately and finally to a combination of the two datasets.

<u>Decision Tree Analysis</u>

Decision tree analysis, or classification or regression trees (CARTs), is used to

predict the membership of individuals into one population group based on predictor

variables (i.e., macromorphoscopic traits/craniometric variables). The tree-like diagrams

are generated from the root node (the top of the tree) to form accompanying branches and

leaves (Williams 2011). Based on decisions or answers to conditional questions or tests,

the direction the node goes and the direction the branches split is determined (see Figure

2*). These branches may continue to branch into other nodes based on additional

questions or tests that will determine if another split will be made or if it will end in a leaf

node that is unable to be split further (i.e. terminal node). These leaf nodes contain the

decisions to these questions/tests.

For example, in Figure 2, the decision tree begins with the macromorphoscopic

trait, inferior nasal aperture (INA). If the INA score is 1, then the individual is classified

as Black. If the INA score is 2 or greater, the next trait to be evaluated is nasal aperture

28

width (NAW). After nasal aperture width (NAW) is evaluated, subsequent traits such as inferior nasal aperture (INA), anterior nasal spine (ANS), inter-orbital breadth (IOB), and nasal bone shape (NBS) are assessed until the leaf nodes are terminated and cannot be further split. Once a terminal node is reached, the individual is classified into a group.
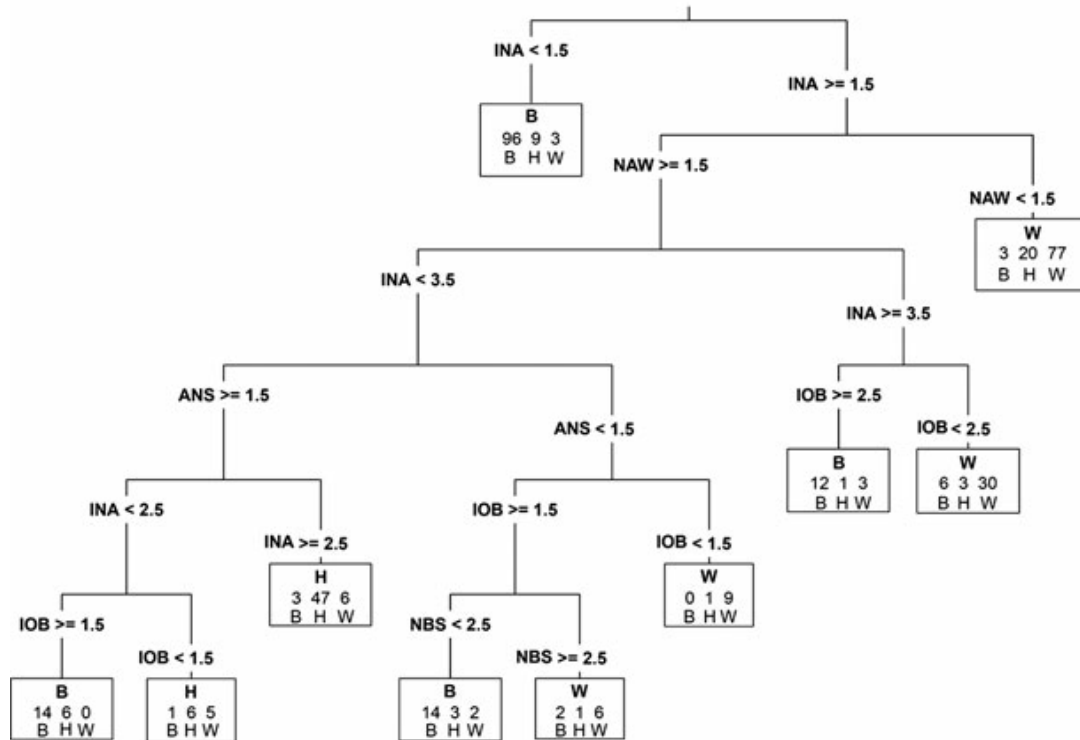
Figure 2. Graphic of decision tree using macromorphoscopic traits for ancestry estimation (Hefner and Ousley 2014)

Random Forest Modeling

Random forest modeling is a dating mining technique for classification and/or regression. Random forest modeling is a classification algorithm composed of a series of individual classification trees (Breiman 2004). To implement random forest modeling, a series of decision trees are generated (normally between 100-500 trees) and each tree is

based on a random selection of observations and predictor variables (Williams 2011).

Each tree on its own makes a prediction based on class membership and those predictions

contribute to the overall random forest prediction (StatSoft 2014). For classification

purposes, majority voting is used to assign the unknown. In other words, if the majority

of the trees classify an individual into a particular group, then the overall random forest

classification follows suit according to that majority vote (Williams 2011). For

regression purposes, the decisions made across all trees in the forest are averaged

(StatSoft 2014).

Each classification tree within the total forest is grown through the introduction of

random vectors influencing the growth of the tree (Breiman 2001). The random forest

algorithm randomly selects the individuals and the variables to be considered (Williams

2011). In this way, the randomness alleviates any potential outliers or potential issues of

overfitting the data to the model. To begin building a random forest model, bagging or

bootstrap aggregation is used to create multiple decision trees by creating separate bags

of randomized observations from the original data into a training data set (a percentage of

the original data "set aside" to serve as training data for the model construction). Single

observations may be used in multiple bags, and normally the training set represents

approximately 2/3 of the original sample. The observations left over may be used for as a

"test" sample to further validate the model.

There is an additional level of randomness when building a random forest model

by the selection of variables at each stage of a split within the decision tree. For every

node in the model, a select number of variables are sampled between those observations a

decision, or split, is assessed. If the model using that split performs better than other

random splits, the model is included in the final forest. If not, another set of random variables are assessed.

Artificial Neural Networks

Artificial neural networks (aNNs) are non-parametric techniques that explore the relationship of data using a model analogous to "biological neural networks," which also stem from Artificial Intelligence research (Williams 2011). The neural networks generated by the user take the gathered data using training algorithms to understand the framework of the data (StatSoft 2014). Complex relationships between the data inputs and outputs are assessed. The data input is represented by the original data while the outputs represent the predicted outcomes of that model.

To begin, the data must be gathered to create a training set from a series of variables that are influential (StatSoft 2014). The model receives inputs from data or from other neurons within the network with their own magnitude, which passes through an "activation function" resulting in outputs. Artificial neural networks are used to establish the probability of group membership to a particular class. The relationship between the input and output data is explored throughout its application. The complexity of aNNs make them intuitively difficult to explain (hence the term "Black Box" methods). Rigorous cross-validation and appropriate sample sizes are generally necessary to avoid overfitting issues.

Support Vector Machines

Support vector analysis (also a Black Box method) creates a model using support vectors to identify a hyperplane which best separates groups. These models use a priori

membership in the algorithm (Williams 2011) to identify the individuals along the outer

edges of multivariate space that divide into various group membership (StatSoft 2014).

This is quite different from other classification methods that utilize distance measures to

the centroid (DFA) or grand mean structures (GM methods). Normally, a maximum

margin or space dividing the groups is desirable (StatSoft 2014), however, SVM fits the

hyperplane to minimize the distances between the line and the outlaying individuals (i.e.,

the support vectors). Typically with the introduction of more than two input variables,

the dividing hyperplane may never be a straight line, however, by instituting a "kernel

trick" to fit a curvilinear line through the data points, the data can be transformed to a

linear feature space (Williams 2011). The data may be separated by curves and

additional techniques will be employed to create linear lines of separation known as

hyperplane classifiers that are commonly executed by support vector machines (StatSoft

2014) (see Figure 3).
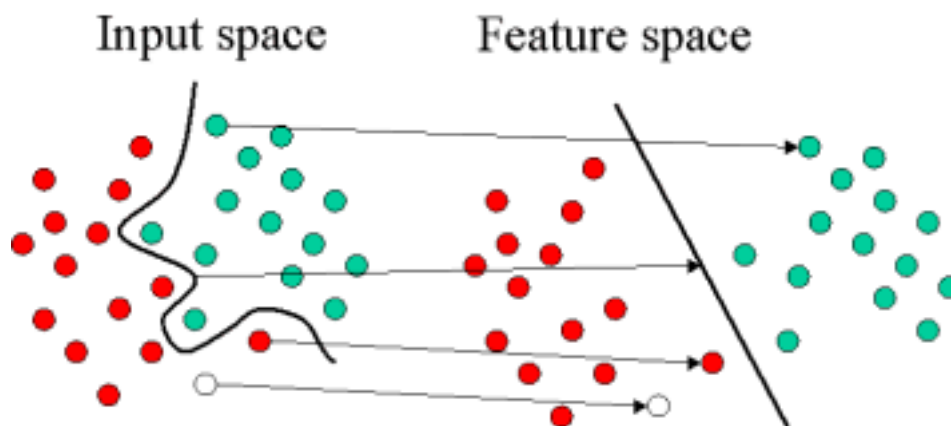
Figure 3. Mapped data with hyperplane using support vector machines (taken from Statistica textbook chapter - Support Vector Machines Overview 2014:1)

III. RESULTS

Trait Frequency Distributions

Trait frequency distributions were calculated by applying cross-tabulations using the SPSS Statistics Package 22.0.0. The frequency distributions for each of the eleven traits are shown below for each of the three population groups (see Table 6).

Table 6. Frequency distribution of macromorphoscopic traits in three populations

Character State													
Anterior Nasal Spine	**Blank**		**1**		**2**		**3**						**Total**
	n	%	n	%	n	%	n	%					
Black	1	2.8	26	72.2	8	22.2	1	2.8					36
White	4	4.1	13	13.4	57	58.8	23	23.7					97
Hispanic	4	8.5	19	40.4	18	38.3	6	12.8					47
Inferior Nasal Aperture	**Blank**		**1**		**2**		**3**		**4**		**5**		**Total**
	n	%	n	%	n	%	n	%	n	%	n	%	
Black	0	0.0	0	0.0	1	2.8	17	47.2	16	44.4	1	2.8	36
White	0	0.0	1	1.0	4	4.1	81	83.5	11	11.3	4	4.1	97
Hispanic	1	2.1	0	0.0	9	19.1	34	72.3	2	4.3	4	8.5	47
Interorbital Breadth	**Blank**		**1**		**2**		**3**						**Total**
	n	%	n	%	n	%	n	%					
Black	0	0.0	14	38.9	17	47.2	5	13.9					36
White	0	0.0	62	63.9	31	32.0	4	4.1					97
Hispanic	0	0.0	22	46.8	23	48.9	2	4.3					47
Malar Tubercle	**Blank**		**1**		**2**		**3**						**Total**
	n	%	n	%	n	%	n	%					
Black	0	0.0	15	41.7	21	58.3	0	0.0					36
White	0	0.0	72	74.2	23	23.7	2	2.1					97
Hispanic	1	2.1	22	46.8	21	36.1	3	6.4					47
Nasal Aperture Width	**Blank**		**1**		**2**		**3**						**Total**
	n	%	n	%	n	%	n	%					
Black	0	0.0	1	2.8	29	80.6	6	16.7					36
White	1	1.0	49	50.5	46	47.4	1	1.0					97
Hispanic	2	4.3	13	27.7	23	48.9	9	19.1					47

Table 6 -Continued Frequency distributions of macromorphoscopic traits in three populations

Nasal Bone Contour	Blank		0		1		2		3		4		5		Total
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	
Black	1	2.8	13	36.1	10	27.8	0	0	9	25.0	1	2.8	1	2.8	36
White	1	2.8	11	11.3	65	67.0	0	0	16	16.5	4	4.1	4	4.1	97
Hispanic	1	2.8	12	25.5	22	46.8	1	2.1	8	17.0	4	8.5	4	8.5	47

Nasal Bone Shape	Blank		1		2		3		4		Total
	n	%	n	%	n	%	n	%	n	%	
Black	1	2.8	8	22.2	18	50.0	5	13.9	4	11.1	36
White	1	1.0	8	8.2	67	69.1	19	19.6	2	2.1	97
Hispanic	6	12.8	6	12.8	24	51.1	11	23.4	0	0.0	47

Nasal Overgrowth	Blank		0		1		Total
	n	%	n	%	n	%	
Black	0	0.0	21	58.3	15	41.7	36
White	1	1.0	64	66.0	32	33.0	97
Hispanic	6	12.8	16	34.0	25	53.2	47

Orbital Shape	Blank		1		2		3		Total
	n	%	n	%	n	%	n	%	
Black	0	0.0	17	47.2	18	50.0	1	2.8	36
White	0	0.0	40	41.2	48	49.5	9	9.3	97
Hispanic	1	2.1	10	21.3	32	68.1	4	8.5	47

Transverse Palatine Suture	Blank		0		1		2		3		Total
	n	%	n	%	n	%	n	%	n	%	
Black	6	16.7	0	0.0	1	2.8	19	52.8	10	27.8	36
White	12	12.4	6	6.2	16	16.5	44	45.4	19	19.6	97
Hispanic	9	19.1	3	6.4	14	29.8	18	38.3	3	6.4	47

Palate Shape	Blank		1		2		3		4		Total
	n	%	n	%	n	%	n	%	n	%	
Black	0	0.0	10	27.8	6	16.7	9	25.0	11	30.6	36
White	6	6.2	36	37.1	24	24.7	16	16.5	15	15.5	97
Hispanic	2	4.3	7	14.9	18	38.3	13	27.7	7	14.9	47

Twenty-four crania from the University of Tennessee – William M. Bass Donated Skeletal Collection were evaluated for interobserver reliability using Fleiss' kappa (see Table 7). These crania were the only individuals sampled by both the author and Hefner (2009). Only macromophoscopic data was used.

Table 7. Results of the Fleiss' kappa interobserver reliability analysis for macromorphoscopic traits

Trait	Interobserver Error, Fleiss' Kappa *k*	Interpretation of Fleiss' Kappa
ANS	0.193	Slight agreement
INA	0.279	Fair agreement
IOB	0.202	Slight agreement
MT	0.262	Fair agreement
NAW	0.509	Moderate agreement
NBS	-0.043	Poor agreement
Novg	0.583	Moderate agreement

Decision Tree Analysis for Macromorphoscopic Dataset

For the decision tree analysis, the classification begin at the top of the tree with ANS (anterior nasal spine) (see Figure 4). If the score for ANS is l then the individual is

classfied as Black. Next, INA (inferior nasal aperture) is evaluated. If INA is greater

than 4, then the individual is classified as Black. If the INA is 3 or less, then the

individual is classified as Hispanic. OS, orbital shape is also evaluated. If OS is 2 or

greater, the individual is Hispanic. If orbital shape is 1, then the individual is White.

   If ANS is greater than 1 then the individual is classified as White and then NAW

(nasal aperture width) is evaluated. If NAW is 2 or more, then the individual is classified

as White and MT (malar tubercle) is consulted until they end into terminal nodes of

White or Hispanic. If NAW has a score of 1, then the individual is classified as White.
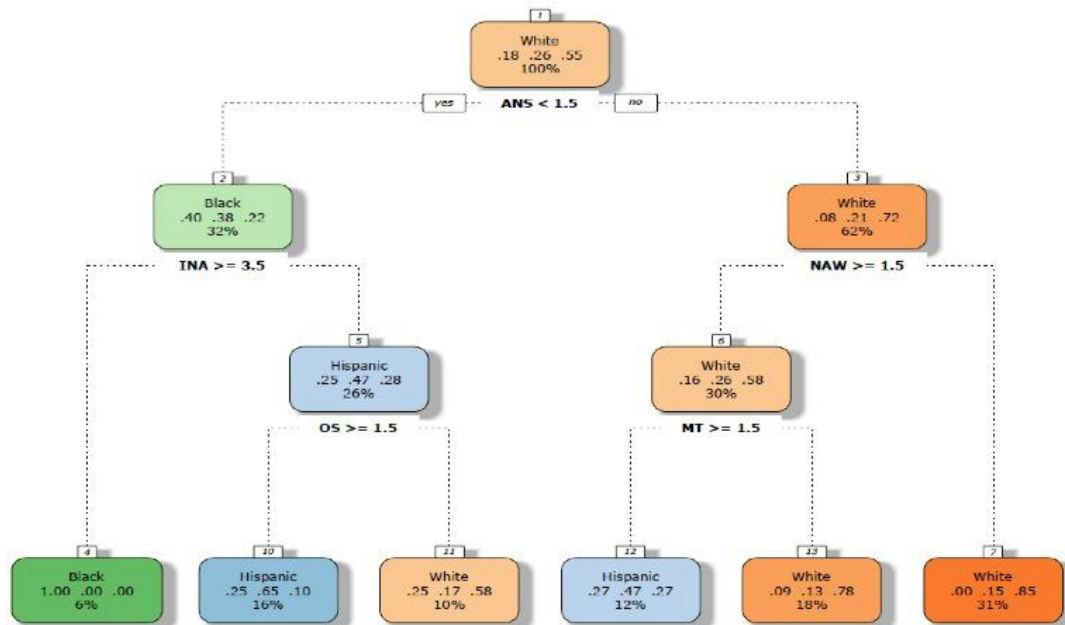
Figure 4. Graphic of the decision tree analysis for the macromorphoscopic dataset (decision tree graphic generated from application of analysis)

Random Forest Model for the Macromorphoscopic Dataset

The random forest model for the macromorphoscopic dataset identified the importance of all 11 variables for the overall classification based on two random forest model measures of variable importance (see Figure 5). The macromorphoscopic traits of the greatest importance were anterior nasal spine (ANS). Inferior nasal aperture (INA), nasal aperture width (NAW), orbital shape (OS), and nasal bone contour (NBC). The macromorphoscopic traits of lesser importance were nasal bone shape (NBS) and nasal overgrowth (Novg).
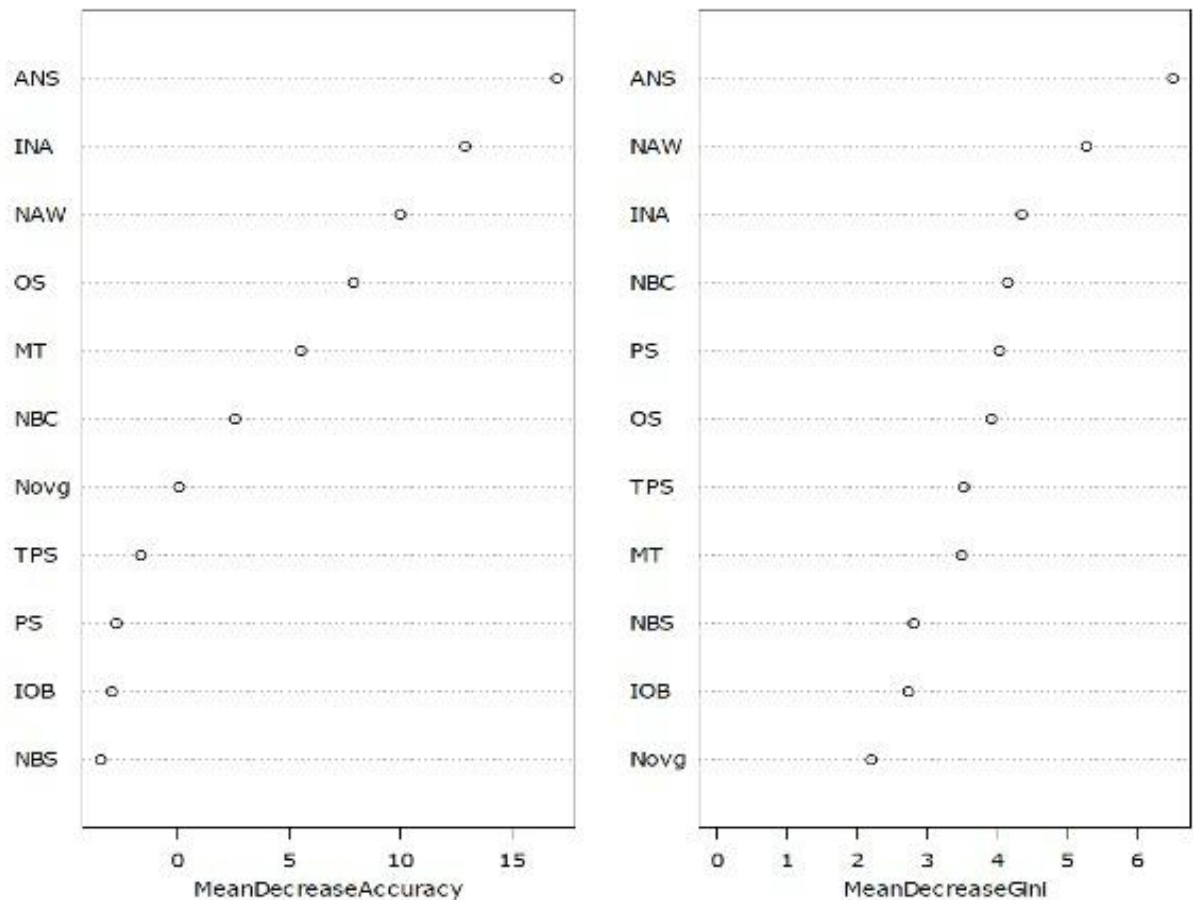
Figure 5. Variable importance of the macromorphoscopic traits in the random forest model

Combined Dataset for the Random Forest Model

The random forest model for the combined dataset of both macromorphoscopic traits and craniometric variables identified variables based on their importance to the overall classification based on two random forest model measures of variable importance (see Figures 6-7). Mostly craniometric variables were identified as having the most importance. The craniometric variable of the greatest importance was palate breadth,

39

external (MAB). Craniometric variables such as NDA, PRA, bimaxillary breadth

(ZMB), and subspinale radius (SSR) were also important contributing variables. The

macromorphoscopic trait that was identified as most important was anterior nasal spine

(ANS). The traits of lesser importance were the macromorphoscopic trait, IOB

(interorbital breadth) and the craniometric variable, nasion subtense fraction (FRF).
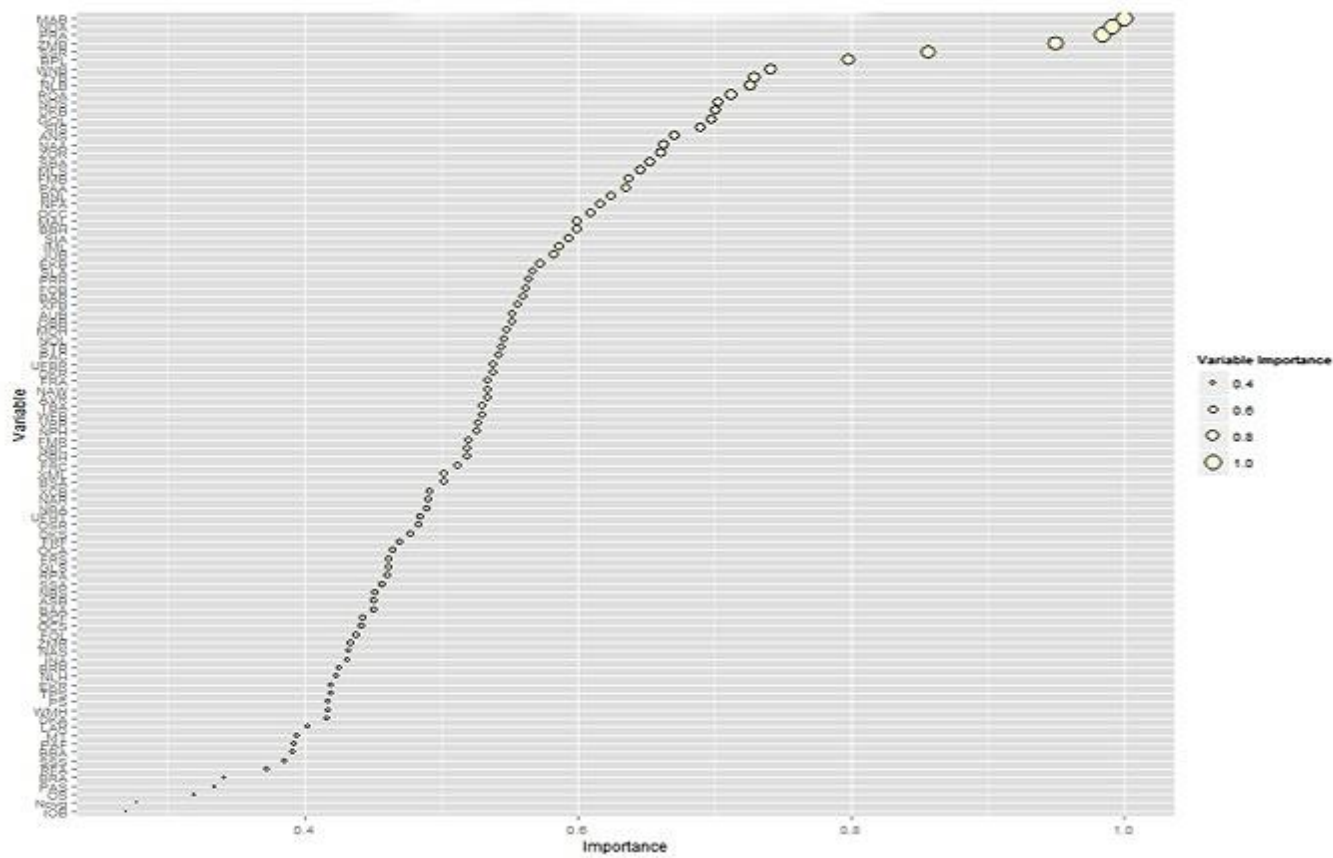
Figure 6. Variable importance of the combined dataset in the random forest model (plot)

41
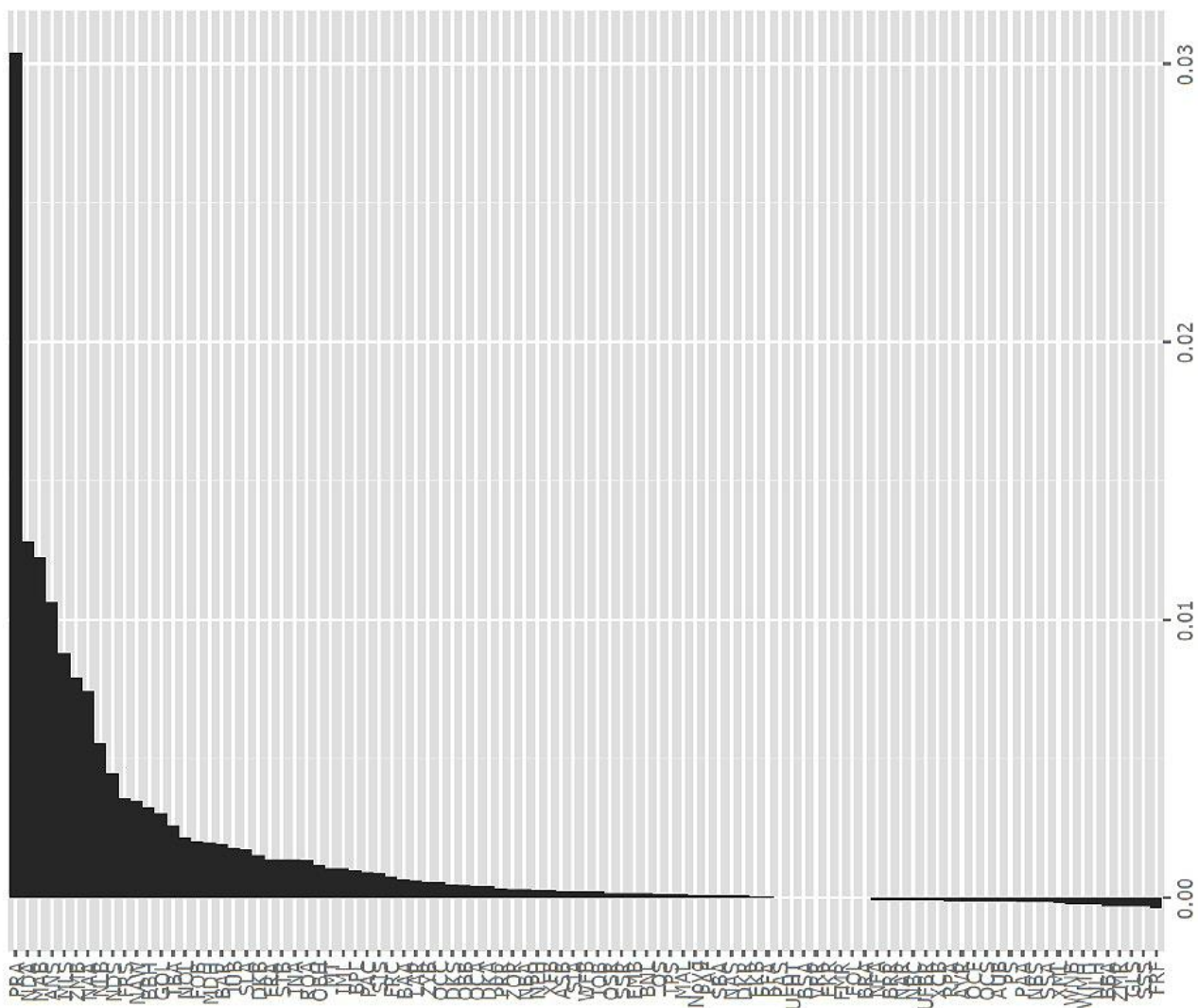
Figure 7. Variable importance of the combined dataset in the random forest model (bar)

Summarization of the Classification Accuracies of the Five Methods for the
Macromorphoscopic, Metric , and Combined Datasets

A cross comparison of the classification accuracies of the five methods for the

macromorphoscopic, metric, and combined datasets is summarized (see Table 8). The

highest overall classification accuracies for each dataset are highlighted in the

accompanying table. The random forest model provided the highest overall classification

accuracies for both the metric and macromorphoscopic datasets at 95% and 90%

respectively. For the combined dataset, the support vector machine had the highest

classification accuracy of 90%. Overall, the classification accuracies for each method

were fairly high with the lowest overall being the decision tree method for the

macromorphoscopic dataset (69.2%). The highest overall classification accuracy was the

random forest model for the metric dataset (95%).

**Table 8. Overall classification accuracies in three-group analyses using metric,
macromorphoscopic, and combined datasets**

Method	Metric	Macromorphoscopic	Combined
Decision Tree	77.09	69.2	77.09
Random Forest	95*	90*	87.75
Support Vector Machine	92.5	70	90*
Artificial Neural Network	93.75	78.62	85.18
Linear Discriminant Function	88	-	-

* Denotes highest percentages of classification for each dataset.

Analysis of the Macromorphoscopic Dataset

The classification accuracies for the macromorphoscopic dataset using a three-

group analysis for the four machine learning methods are summarized (see Table 9). The

artificial neural network (aNN) had a high probability of correctly classifying the individuals in the Black group at 73.9%. For the Hispanic group, the random forest model (RFM) had the highest classification accuracy of 87%. Two methods (random forest model and support vector machine) classified the White group with the highest classification accuracies of 100%. Of the four methods, the decision tree analysis had the lowest classification accuracy for both the Black and Hispanic groups. The artificial neural network had the lowest classification accuracy for the White group.

Table 9. Classification accuracies for macromorphoscopic analysis of machine learning methods

	Decision Tree	Random Forest Model	Support Vector Machine	Artificial Neural Network
Black	38.9	83.3	50	73.9*
Hispanic	55.3	87*	60.9	78.2
White	87.5	100*	100*	85.7

*Denote highest percentages of classification for each population group.

Analysis of the Craniometric Dataset

The classification accuracies for the craniometric dataset using a three-group analysis for the different methods are summarized (see Table 10). Specific methods (random forest model, support vector machine, artificial neural network, and linear discriminant function) all have high classification accuracies for the different population groups. The artificial neural network (aNN) and the linear discriminant function analysis work best at classifying the Black group correctly (100%). For the Hispanic group, the support vector machine provides high classification accuracies of 100%. Similarly, the random forest model provides high classification accuracies of 100% for the White group. Of the methods applied, the decision tree analysis had low classification accuracies for the Black and Hispanic groups at 77.8% and 57.4% respectively. The

linear discriminant function analysis had a lower classification accuracy for the White group at 84.6%.

Table 10. Classification accuracies for craniometric analysis of machine learning methods

	Decision Tree	Random Forest Model	Support Vector Machine	Artificial Neural Network	Linear Discriminant Function
Black	77.8	83.3	83.3	100*	100*
Hispanic	57.4	95.7	100*	83.3	85.7
White	86.5	100*	81.8	96.3	84.6

*Denotes highest percentages of classification for each population group.

Analysis of the Combined Dataset

The classification accuracies for the combined dataset using a three-group analysis for the different methods are summarized (see Table 11). Specific methods (random forest model, support vector machine, artificial neural network) all have high classification accuracies for the different population groups. The artificial neural network (aNN) work best at classifying the Black group correctly (85.7%). For the Hispanic group, the support vector machine provides high classification accuracies of 100%. Similarly, the random forest model provides high classification accuracies of 97.9% for the White group. Of the methods applied, the support vector machine had low classification accuracies for the Black and the White groups at 66.7% and 81.8% respectively. The decision tree analysis had low classification accuracies for the Hispanic group at 57.4%.

Table 11. Classification accuracies for combined analysis of machine learning methods

	Decision Tree	Random Forest Model	Support Vector Machine	Artificial Neural Network
Black	77.8	72.2	66.7	85.7*
Hispanic	57.4	81.8	100*	86.7
White	86.5	97.9*	81.8*	82.4

*Denotes highest percentages of classification for each population group.

IV. DISCUSSION

To assess ancestry, macromorphoscopic or metric analysis is conducted. Generally, research involving macromorphoscopic trait expression can be effective for ancestry estimation when it is accompanied by robust statistical analyses and known classification accuracies (Hefner 2009; Hefner and Ousley 2014; Hefner et al. 2014; Klales and Kenyhercz 2014). In previous literature, qualitative analysis has been geared more towards the typological approach as these macromorphoscopic traits are assessed visually. However, current literature is constantly evolving by incorporating both non-metric traits strengthened with statistical analyses (Hurst 2012; Hefner 2009; Hefner and Ousley 2014; Hefner et al 2014). Metric analysis for ancestry estimation normally involves standardized measurements of the crania and is subjected to statistical analysis to classify an individual. With the addition of a variety of reference groups for comparison, craniometric analysis has less opportunity for subjectivity when rigorous statistical analysis may be applied. The analysis of macromorphoscopic data needs to be strengthened to that same level of analysis. The addition of machine learning techniques provides another dimension for interpreting the data in an unconventional manner (Hefner and Ousley 2014; Hefner et al. 2014). Uncovering and interpreting the patterns that emerge through machine learning methods is another aspect of ancestry estimation that needs to be further developed. Exploring the significance of how and to what degree the variety of macromorphoscopic and craniometric methods contribute to ancestry estimation and which traits specifically are most important for ancestry estimation was one of the most important questions posited within this research.

Trait Frequency Comparison with Hefner (2009)

Frequencies of corresponding non-metric traits in both Hefner (2009) and this study were compared further. The non-metric traits from this study were taken from the Hefner (2009) study. It is important to note that the American White and American Black samples in the Hefner study are considerably larger than those in the present study. Therefore, this variable may be a contributing factor to the differences in reported frequencies. Only groups that were evaluated in both studies were used for the comparison. Therefore, trait frequencies of the White and Black population groups were utilized. Anterior nasal spine, inferior nasal aperture, interorbital breadth, malar tubercle, nasal aperture width, nasal bone contour, nasal bone shape, nasal overgrowth, and transverse palatine suture were evaluated.

Anterior nasal spine (Hefner 2009 vs. Wun 2014*)*

For the Black population groups, the slight projecting anterior nasal spine had the most prevalent trait frequencies in both studies. The frequencies were comparable to one another in the sense that both were greater in over 50% of the samples (see Table 12). This character state of anterior nasal spine was probably the easiest to observe. However, at times, the anterior nasal spine may be subject to poor preservation because that area of the face is delicate and the thin spine may break. There was some discrepancy with the trait frequencies of the White population group between the two studies. In the Hefner (2009) sample, the trait frequencies were somewhat close between a score of 1 (slight projection) and 3 (marked projection) with frequencies of 36.3% and 37.7% respectively. The present study found that the character state of intermediate projection was most prevalent with a frequency of 58.8%. The subjective nature of evaluating anterior nasal

spine comes into play when evaluating scoring between two observers. There is not

quantitative manner of measuring the transition from intermediate to marked projection.

Table 12. Comparison of the frequency distributions for anterior nasal spine (ANS) Hefner 2009 vs. Wun 2014

	Hefner 2009				Wun 2014			
	Black		White		Black		White	
ANS	n	%	n	%	n	%	n	%
Blank	0	0.00	0	0.00	1	2.80	4	4.10
1	152	69.7	53	36.3	26	72.2	13	13.4
2	44	20.2	38	26.0	8	22.2	57	58.8
3	22	10.1	55	37.7	1	2.8	23	23.7

Inferior nasal aperture (Hefner 2009 vs. Wun 2014)

The trait frequencies for inferior nasal aperture were considerably different

between the Hefner (2009) study and this present study (see Table 13). Within the Black

sample in Hefner's study, a score of 2 appeared in about 28.9% of the sample while a

score of 4 was most prevalent in the White sample (2009). In the present study, a score

of 3 appeared in approximately 47.2% of the Black sample and in approximately 83.5%

of the White sample. The significant differences in trait frequencies could be explained

by two factors. First, the large sample size in Hefner's study could alter the various trait

frequencies. Second, in Hefner's interobserver reliability analysis, inferior nasal aperture

was a trait only moderately agreed upon (2009). However, the trait frequencies for a

score of 1-2 were the two character states that were closest between the two studies.

Therefore, this could explain the differences in scoring and trait frequencies.

Table 13. Comparison of frequency distributions for inferior nasal aperture (INA) Hefner 2009 vs. Wun 2014

	Hefner 2009				Wun 2014			
	Black		White		Black		White	
INA	n	%	n	%	n	%	n	%
Blank	0	0.00	0	0.00	0	0.00	0	0.00
1	64	29.4	1	0.7	0	0.0	1	1.0
2	63	28.9	5	3.4	1	2.8	4	4.1
3	47	21.6	35	24.0	17	47.2	81	83.5
4	29	13.3	60	41.1	16	44.4	11	11.3
5	15	6.9	45	30.8	1	2.8	4	4.1

Interorbital breadth (Hefner 2009 vs. Wun 2014)

The next trait that was evaluated between the two studies was interorbital breadth (see Table 14). The scores differed between the two studies. A score of 3 (broad) was present in 56% of the Black sample and a score of 2 (intermediate) was present in 63% of the White sample in the Hefner study (2009). In the present study, a score of 2 (intermediate) was present in 47.2% of the Black sample, and a score of 1 (narrow) was present in 63.9% of the White Sample. In his study, Hefner used crania from the Robert J. Terry collection for the White sample. Most of the individuals in this sample were collected from 1917-1966 and most represented cadavers that were used for anatomical study at the Washington University at St. Louis medical school (Hunt and Albanese 2005). For the present study, the White sample consisted of individuals from the William M. Bass Donated Skeletal Collection. The population demographics each collection differed by geography and also the type of individuals that made up each collection.

Table 14. Comparison of frequency distributions for interorbital breadth (IOB) Hefner 2009 vs. Wun 2014

	Hefner 2009				Wun 2014			
	Black		White		Black		White	
IOB	n	%	n	%	n	%	n	%
Blank	0	0.00	0	0.00	0	0.00	0	0.00
1	21	9.6	45	30.8	14	38.9	62	63.9
2	75	34.4	92	63.0	17	47.2	31	32.0
3	122	56.0	9	6.2	5	13.9	4	4.1

Malar tubercle (Hefner 2009 vs. Wun 2014)

The frequency distributions for the variable, malar tubercle differed between the White and the Black groups between the two studies (see Table 15). In the Hefner study, the score of 0 was more frequent in the Black sample (50.5%) and a score of 0 was more frequent in the White sample (51.40). In the present study, the score of 2 was more frequent in the Black sample (58.3%) and a score of 1 was more frequent in the White sample (74.2%).

Table 15. Comparison of frequency distributions for malar tubercle (MT) Hefner 2009 vs. Wun 2014

	Hefner 2009				Wun 2014			
	Black		White		Black		White	
MT	n	%	n	%	n	%	n	%
Blank	0	0.00	0	0.00	0	0.00	0	0.00
0	110	50.50	75	51.40	0	0	0	0
1	60	27.5	47	32.2	15	41.70	72	74.20
2	32	14.7	18	12.3	21	58.3	23	23.7
3	16	7.3	6	41.1	0	0.0	2	2.1

Nasal aperture width (Hefner 2009 vs. Wun 2014)

The frequency distributions of the nasal aperture width between the Hefner (2009) study and the present study were evaluated (see Table 16). In the Hefner study, a score of 3 was most frequently expressed in the Black sample (55.5%), and a score of 1 was most

frequently expressed in the White sample (54.1%). To contrast, in the present study, a

score of 2 was most frequently expressed in the Black sample (80.6%), and a score of 1

was most frequently expressed in the White sample (50.5%). Although the score of 1

was the highest frequency expressed, the score of 2 (47.4%) was a close second.

Therefore, a single character state did not encompass the expression of nasal aperture

width within the White sample. The three character states within this trait can be highly

variable because of their subjective nature.

Table 16. Comparison of frequency distributions for nasal aperture width (NAW) Hefner 2009 vs. Wun 2014

| | Hefner 2009 | | | | Wun 2014 | | | |
| | Black | | White | | Black | | White | |
NAW	n	%	n	%	n	%	n	%
Blank	0	0.00	0	0.00	0	0.00	1	1.00
1	8	3.7	79	54.1	1	2.8	49	50.5
2	89	40.8	48	32.9	29	80.6	46	47.4
3	121	55.5	19	13.1	6	16.7	1	1.0

Nasal bone shape (Hefner 2009 vs. Wun 2014)

The trait frequencies of the Hefner (2009) sample was greatly distributed amongst

the five character states, while in the present study, the high frequencies were relegated to

specific character states (see Table 17). In the Hefner study, a score of 0 was expressed

with the highest frequency in the Black sample (52.3%), while a score of 3 was expressed

with the highest frequency in the White sample (25.3%). In the present study, a score of

2 was expressed with the highest frequency in the Black sample (50%) and the White

sample (69.1%). A score of 0 was mostly recorded in the Hefner sample, however not in

the present study at all. Perhaps, there may have been a misunderstanding about the trait

definitions that could have affected the observations.

**Table 17. Comparison of frequency distributions for nasal bone shape (NBS)
Hefner 2009 vs. Wun 2014**

	Hefner 2009				Wun 2014			
	Black		White		Black		White	
NBS	n	%	n	%	n	%	n	%
Blank	0	0.00	0	0.00	1	2.80	1	1.00
0	114	52.30	11	7.5	0	0	0	0
1	50	22.9	23	15.8	8	22.2	8	8.2
2	22	10.1	27	18.5	18	50.0	67	69.1
3	23	10.6	37	25.3	5	13.9	19	19.6
4	9	4.1	48	32.9	4	11.1	2	2.1

Nasal overgrowth (Hefner 2009 vs. Wun 2014)

        The presence and absence of nasal overgrowth between the Hefner (2009) study

and the present study was evaluated (see Table 18). Since nasal overgrowth was an

important feature highlighted in both Rhine (1990) and Hurst (2012) typically expressed

in Hispanics, the trait frequencies of the White and Black samples were evaluated

carefully to see if nasal overgrowth was expressed in other population groups. In both

studies, the absence of nasal overgrowth was apparent for both the White and the Black

samples. In the Hefner (2009) study, the absence of nasal overgrowth was expressed in

68.1% of the Black sample and 52.7% of the White sample. In the present study, the

absence of nasal overgrowth was expressed in 58.3% of the Black sample and 66% of the

White sample. This is not to say that these population do not exhibit nasal overgrowth,

however a higher percentage of the samples did not.

Table 18. Comparison of frequency distributions for nasal overgrowth (Novg) Hefner 2009 vs. Wun 2014

| | Hefner 2009 | | | | Wun 2014 | | | |
| | Black | | White | | Black | | White | |
Novg	n	%	n	%	n	%	n	%
Blank	0	0.00	0	0.00	0	0.00	1	1.00
0	141	68.1	77	52.7	21	58.3	64	66.0
1	66	31.9	69	49.2	15	41.7	32	33.0

Transverse palatine suture (Hefner 2009 vs. Wun 2014)

Scoring for transverse palatine suture was evaluated between the Hefner (2009) and the present study to examine the expression of trait frequencies between the White and the Black samples (see Table 19). In the Hefner (2009) study, a score of 1 was expressed 47.2% for the Black sample and a score of 2 was expressed 33.8% of the White sample. In the present study, a score of 2 was expressed in 52.8% of the Black sample and 45.4% of the White sample. There was similarity in the two studies in that the last character state (score of 4) was minimally expressed in both samples. The differences in population sizes could possibly contribute to the differing trait frequencies.

Table 19. Comparison of frequency distributions for transverse palatine suture (TPS) Hefner 2009 vs. Wun 2014

| | Hefner 2009 | | | | Wun 2014 | | | |
| | Black | | White | | Black | | White | |
TPS	n	%	n	%	n	%	n	%
Blank	0	0.00	0	0.00	6	16.70	12	12.40
0	33	18.30	42	29.0	0	0.0	6	6.2
1	85	47.2	40	27.6	1	2.8	16	16.5
2	45	25.0	49	33.8	19	52.8	44	45.4
3	17	9.4	14	9.7	10	27.8	19	19.6
4	0	0.0	0	0.0	0	0.0	0	0.0

Trait Frequency Comparison with Hurst (2012)

In the Hurst (2012) study, the author evaluated 26 macromorphoscopic traits adopted from various trait lists (Birkby et al. 2008; Hefner 2009; Gill 1998; Napoli and Birkby 1990; Rhine 1990). Hurst derived the final two traits that possibly contributed to the estimation of ancestry (2012). After applying forward stepwise discriminant function analysis, eight of the 26 traits were found to best differentiate between the three samples including incisor shoveling, anterior malar projection nasal sill, oval window visualization, enamel extensions, anterior nasal spine, nasal aperture width and prognathism (Hurst 2012). Two of these traits were derived from Hefner (2009): anterior nasal spine and nasal aperture width. Contrary to the findings of Hurst (2012), in the present study, nasal overgrowth appeared in approximately 53% of the Hispanic sample, and of the three groups, nasal overgrowth was most prevalent in Hispanics. However, nasal overgrowth was consistently present in both the White (32.7%) and Black (41.7%) groups as well. Overall, the majority of individuals displaying nasal overgrowth were of Hispanic descent.

Based on the macromorphoscopic traits that were evaluated in the present study, they were cross-referenced with the findings of Hurst (2012), describing the trait expressions that were characteristic of Southwest Hispanics. Of the eight traits that were identified as being characteristic of this population group, three were evaluated in the present study. Hurst concluded that individuals of Southwest Hispanic origin expressed a moderate anterior malar projection, intermediate anterior nasal spine, and medium nasal aperture width (2012). In the present study, the Hispanic sample exhibited a frequency of moderate anterior malar projection of 36.1% while a score of 1 (minimal projection) was

expressed in a greater frequency of 46.8% (see Table 20). However, the score of

moderate anterior mal projection was a close second in this study. The frequencies

expressed in Hurst's study concluded that there was a moderate projection in 88.5% of

the Hispanic sample (2012).   There may be a fair amount of variation in trait expression

of the malar tubercle, since this trait can be measured. Discrepancies in recording could

also play a role in affecting the differences in trait frequencies.

Table 20. Comparison of frequency distributions for malar tubercle (MT) Hurst 2012 vs. Wun 2014

	Hurst 2012	**Wun 2014**
	Hispanic	**Hispanic**
MT	**%**	**%**
Blank	0	2.1
1	3.3	46.8
2	88.5	36.1
3	8.2	6.4

Hurst identified moderate anterior nasal spine as another characteristic trait of individuals

of Southwest Hispanic descent (2012). A moderate anterior nasal spine was expressed in

39.6% of the Hispanic sample and a marked anterior nasal spine with a frequency of

37.7% (Hurst 2012) (see Table 21). In the present study, slight anterior nasal spine was

expressed in 40.4% of the Hispanic sample with a moderate anterior nasal spine with a

frequency of 38.3%. The subjectivity of this particular trait could explain the differences

in trait frequencies.

Table 21. Comparison of frequency distributions for anterior nasal spine (ANS) Hurst 2012 vs. Wun 2014

	Hurst 2012	Wun 2014
	Hispanic	Hispanic
ANS	%	%
Blank	0	8.5
1	22.6	40.4
2	39.6	38.3
3	37.7	12.8

Lastly, the trait frequencies of nasal aperture width were evaluated between the two studies (see Table 22). The highest frequencies that were expressed in both Hispanic samples in both studies were a moderate nasal aperture width. The Hurst sample had a trait frequency of 57.4, while the present study had a trait frequency of 48.9 for a score of 2 (moderate nasal aperture width). For the most part, the trait frequencies overall were similar between the Hurst (2012) study and the present study. Agreement of this trait designation in the Hispanic samples may corroborate that this particular character state is consistent with individuals of Hispanic origin. Furthermore, the amount of agreement and close trait frequencies between two observers may suggest that nasal aperture width is a trait that is more easily distinguishable based on its three character states.

Table 22. Comparison of frequency distributions for nasal aperture width (NAW) Hurst 2012 vs. Wun 2014

	Hurst 2012	Wun 2014
	Hispanic	Hispanic
NAW	%	%
Blank	0	4.3
1	27.9	27.7
2	57.4	48.9
3	14.8	19.1

Trait Frequency Comparison with Rhine (1990)

Frequencies of corresponding non-metric traits in both Rhine (1990) and this study were compared further. The traits compiled in Rhine (1990) are situated in the mid-face and many originate from Hooton's Harvard list and have been consistently used by forensic anthropologists for ancestry estimation. Non-metric traits evaluated in both Rhine (1990) and the present study was cross-referenced and the trait frequencies were compared. Orbital shape, nasal opening, nasal overgrowth, nasal spine, malar tubercle, dental arcade shape, and palatine suture were evaluated.

Orbital shape (Rhine 1990 vs. Wun 2014)

Frequencies for orbital shape Reference samples for the Hispanic population were less than five, which would affect the frequencies since the sample sizes were smaller. According to Rhine's evaluation of orbital shape, rounded orbits were most common in the Hispanic group, while sloping orbits were most common in the White population (1990). This was not the case in the present study. The circular (rounded) eye orbit shape was most common amongst all groups evaluated: Black (50%), White (49%), and Hispanic (68.1%). It is important to note that the angled orbit shape was not even close

in frequency to either the rectangular or circular shape. Rhine commented that the angled orbit shape was most common in the Hispanic crania, however due to sample size and the geographic area of the data, this could account for such a difference in frequency distribution.

Nasal opening (Rhine 1990 vs. Wun 2014)

Another trait looked at was nasal opening, which corresponds with nasal aperture width in the present study. Generally, the White group would display a greater frequency of a narrow nasal opening, while the Hispanic group displayed the intermediate size, and the Black group, a wider opening (Rhine 1990). In the present study, the Hispanic and Black groups had an intermediate nasal aperture width. The trait expression for Hispanics coincides with Rhine, however the Black group did not. The White group did display a narrow nasal opening, which corroborates the data from Rhine (1990).

Nasal overgrowth (Rhine 1990 vs. Wun 2014)

Nasal overgrowth is considered one of the important non-metric traits, especially for describing individuals of Hispanic descent (Birkby et al. 2008; Hurst 2012). For the Hispanic samples, the majority of the individuals did not exhibit presence of nasal overgrowth, and in the other groups, the absence of nasal overgrowth was most common (Rhine 1990). In the present study, overwhelmingly, the Hispanic individuals most commonly displayed nasal overgrowth, while the frequency of other groups (White and Black) tended towards the absence of it.

Anterior nasal spine (Rhine 1990 vs. Wun 2014)

In the Rhine study, anterior nasal spine was consolidated to either presence of a small or large spine (1990). Overwhelmingly, each of the groups had a less pronounced, small nasal spine. This could be due to the delicate nature of the region. A broken anterior nasal spine may be smaller to the observer. In the present study, nasal spines that looked to be broken were designated a blank score. Anterior nasal spine was scored on a continuum as opposed to being either small or large. This additional character state could have altered the frequencies as well. Overall smaller sample sizes of the Hispanic and Black groups in the Rhine study may have contributed to the differing frequencies.

Malar tubercle (Rhine 1990 vs. Wun 2014)

Another trait that was compared between the two studies was malar tubercle. The frequencies may not be entirely comparable because the present study characterized scoring by observing the amount of protrusion of the malar tubercle rather than noting the overall presence or absence of protrusion. Malar tubercle was found in half of the individuals in the Hispanic and in all of the Black individuals (Rhine 1990). The presence of the malar tubercle in the Black sample did not necessarily indicate that this trait was most common in this group. The small sample sizes again that exhibited the trait in all three groups could definitely be a contributing factor. The presence and/or absence of malar tubercle do not provide a complete picture of trait expression. In the present study, malar tubercle was found to be a significant discriminating factor between the three population groups studied. The character state of pronounced tubercle did not have a high frequency rate in any of the three populations indicating that the trace and

medium character states were most prevalent rather than the two extreme states.

Therefore, the presence and absence of a slightly pronounced and/or heavily pronounced

malar tubercle does not provide sufficient information regarding variation between the

three populations.

Dental arcade shape (Rhine 1990 vs. Wun 2014)

The elliptic palate shape was most prevalent in each of the three groups within the

Rhine sample (1990). Variability of this trait is not discussed in Rhine, however there is

a general consensus that each group exhibits, to a degree, a high frequency of an elliptic

palate shape. Gill (1998) asserts that each of the three population groups typically exhibit

a specific, expected palate shape. Gill comments that Whites tend to show a parabolic

palate and Black with a hyperbolic palate shape (1998). He discusses palate shape

frequencies based on various studies (Chapman 1991; Gill and Chapman nd; Gill 1998).

Of the three common types of palate shapes: elliptic hyperbolic, and parabolic, White

populations exhibit a high frequency of the parabolic shape (91%) and Black populations

had high frequencies of both hyperbolic and parabolic palate shapes (46% and 54%

respectively) (Chapman 1991; Gill and Chapman nd; Gill 1998). This analysis goes

against the palate shape trait frequencies of the Rhine (1990) study. Gill asserts that

American Indian populations (N.W. Plains and Peruvian samples) exhibit the elliptic

palate shape in higher frequencies than White and Black populations. In the present

study, the type of plate shape with the highest frequency differed between the three

groups that were evaluated. The group that showed the highest frequency of the elliptic

palate shape was the White population, followed by the Hispanic group with the highest

frequencies of both states of the parabolic shape, and lastly the Black population

exhibited the highest frequencies of the hyperbolic palate shape. The frequencies of the

White group with elliptic palate shape corresponds with the findings of Rhine while the

frequencies of the Black group with hyperbolic palate shape supports the frequencies

reported by Gill (1998), however not as high. The variable palate shape was not highly

significant in discriminating between the three groups within the discriminant function

analysis, however the quantification of palate shape expressions supports the need for

further research in this area.

Furthermore, the introduction of palate shape to the list of macromorphoscopic

traits in this study was significant. Trait frequency data of palate shape amongst the

American White, American Black, and Hispanic sample groups in this study is useful for

future comparison studies related to ancestry estimation. The inclusion of palate shape to

the Macromorphoscopic module would contribute to the further understanding of this

trait from a morphological standpoint. By exploring the macromorphoscopic expression

of palate shape compared to palate shape measurements, the understanding of how

variations in shape and size of palate significantly differ between populations.

Additional palate shape data may be collected for various population groups in the

expansion of knowledge for this particular trait, especially in the case of the Hispanic

sample included within this study.

Assessing Interobserver Reliability

Inter-observer reliability showed that the there was a slight-moderate agreement

between the crania that were sampled by the author and in Hefner (2009). Nasal aperture

width and nasal overgrowth were the two traits that had moderate agreement. This could

be due to the fact that nasal growth is based on presence and absence while nasal aperture

width is scored based on three categories: narrow, medium, and broad. Both of these

traits had character states that of three or less. Therefore it may be easier to score those

traits compared to the others. However at the same time, there may be difficulty due to

the subjectivity of the trait. Therefore, the incorporation of additional observers should

be applied in future research. Anterior nasal spine and interorbital breadth both had fair

agreement. In Hefner (2009), the traits that had the least amount of inter-observer

reliability were nasal bone contour and postbregmatic depression. In this study,

postbregmatic depression was not included as one of the traits that were analyzed. For

future research, the remaining macromorphoscopic traits in Hefner (2009) that were not

included in this study should be used to score the same Hispanic sample and other

populations to generate current frequencies of traits expression.

　　　　Additionally education and experience may also contribute to the slight to

moderate agreement. The author spent a short time reviewing the macromorphoscopic

traits and assessing thirty crania prior to data collection. In contrast, Hefner had more

experience looking at these macromorphoscopic traits on a variety of crania and perhaps

had a more complete experience and confidence that can be attained through constant

assessment and opportunity of viewing these traits.

Assessment of Machine Learning Analysis

　　　　The application of machine learning techniques (Hefner and Ousley 2014;

Williams 2011) in the present study resulted in high classification accuracies for a variety

of those methods. Therefore, the high classification accuracies support other studies that have applied these types of analyses to issues in forensic anthropology (Hefner and Ousley 2014; Hefner et al. 2014; Navega et al. 2013; Navega et al. 2014). In the present study, overall, the classification accuracies performed well with most methods ranging from 80-100%. Based on overall classification accuracies using all three datasets, the highest classification accuracy at 95% applied the random forest model to the metric dataset. The second highest classification accuracy across all three datasets employed the artificial neural network for the metric dataset at 93.75%. The third highest overall classification accuracy was at 92.5% using the support vector machine for the metric dataset.

Generally, it seems that the machine learning methods produced the highest classification accuracies with metric data. This type of data may be better suited for these types of method as opposed to the scoring of the macromorphoscopic data. This is not surprising because the craniometric data was composed of standardized measurements that were instrumentally determined rather than the macromorphoscopic traits that were visually assessed on an ordinal scale by the observer. However, applying machine learning techniques to the macromorphoscopic data produced high classification accuracies of 90% for the random forest model. These high classification accuracies suggested that the random forest model is the best overall machine learning method.

The random forest analysis is particularly helpful because it provided a ranking of variable importance of the predictor variables used for classification. These individual variables that are isolated may be helpful to discover which traits are better at making distinctions between various population groups. In the macromorphoscopic dataset, the

three variables identified were anterior nasal spine (ANS), inferior nasal aperture (INA), and nasal aperture width (NAW). Within the combined dataset (both macro and metric data) using the random forest model, the first few variables of greater importance were craniometric: MAB (palate breadth, external), NDA, and PRA followed by a few macromorphoscopic traits that were significant such as anterior nasal spine and nasal aperture width.

Looking specifically at the classification accuracies of the methods used in the combined dataset, the support vector machine produced the highest classification accuracy of 90%. The classification accuracies in the combined dataset were not as high as the individual datasets. Therefore, combining macromorphoscopic and metric data in a single analysis may not be the most effective at obtaining high classification accuracies for all machine-learning methods. However, the highest classification accuracy using the combined dataset applied support vector machine analysis had a correct classification percentage of 90%.   Further research applying support vector machine analysis may corroborate the utility of this method for using combined macromorphoscopic and metric datasets in future studies since most other machine learning techniques displayed lower classification accuracies.

Looking at each dataset individually, the macromorphoscopic analysis using machine learning methods overall, were generally high. The decision tree analysis had the lowest overall classification accuracy for the Black group with 38.9% being correctly classified. However, the random forest model provided high classification accuracies across all three groups, which supports the idea that this machine learning method may be most useful for analyzing macromorphoscopic data. The support vector machine also

provided 100% correct classification for the White group, as did the random forest

model. In this scenario, the very high classification accuracy meant that the White group

was easier to parse out and separate from the Black and Hispanic groups and may have

overlapping size and shape similarities. It is clear that no one method can easily separate

all three populations into distinct groups.

The machine learning methods in the craniometric dataset performed extremely

well with 100% correct classifications for each group for certain methods. Within the

Black group, the artificial neural network (aNN) and linear discriminant function analysis

had classification accuracies of 100%. For the Hispanic group, the support vector

machine analysis garnered a 100% correct classification and for the White group, the

random forest model successfully classified 100% of the individuals in the sample.

Although a different machine-learning model worked the best for each group, the

craniometric dataset had the best results across the board.

For the combined dataset, the highest classification accuracy for the Black group

was using the artificial neural network (aNN) at 85.7%. This is significantly lower than

the highest classification accuracies for the Hispanic and the White groups. For the

Hispanic group, the support vector machine method had the highest classification

accuracy of 100%. The White group had a classification accuracy of 97.9% for the

random forest model. Generally, the percentages for combined analysis were lower than

the other datasets. Therefore, the machine learning method can be better applied to

individual datasets to obtain the highest classification accuracies. However, many of the

classification accuracies are not considerably low, but fall within the range of 50-86%.

Further research should be conducted to examine the range of variation of the

classification accuracies between individual populations using individual macro and metric datasets for the different machine-learning methods.

Hefner and Ousley (2014) applied similar machine learning methods to estimate ancestry using macromorphoscopic data for White, Black, and Hispanic groups. Therefore, it would be helpful to compare the overall classification accuracies of this study with their own to evaluate any differences or similarities between percentages of correct classification of group membership. Of the machine learning methods, neural network analysis provided the highest overall classification accuracy between all the methods. The top three methods that obtained the highest classification accuracies were neural network analysis (87.8%), support vector machine (86.4%), and random forest (85.5%) (Hefner and Ousley 2014). These three methods were also the top performing techniques in the present study based on the macromorphoscopic dataset. However, the top performing technique was the random forest model with a high classification accuracy of 90% compared to the Hefner and Ousley study with a classification accuracy of 85.5%. However, both studies support the idea that the neural network analysis, support vector machine, and random forest modeling all positively contribute to analyzing macromorphoscopic data individually.

Hefner and colleagues examined macromorphoscopic and craniometric data using both linear discriminant functions and random forest modeling (2014). Discriminant function analyses were applied individually to the macromorphoscopic, craniometric, and combined (macro and metric) datasets while the random forest model was only applied to the combined (macro and metric) dataset (Hefner et al. 2014). The random forest model had a classification accuracy of 89.6 % (Hefner et al. 2014) compared to the present

study that had an overall classification accuracy of 87.75%. The slighter lower classification accuracy in the present study may be influenced by the differences in number of predictor values utilized compared to the study by Hefner and colleagues (2014). Both used the same number of macromorphoscopic traits, but a greater number of craniometric variables were used in the present study. However the classification accuracies for both are similar and support the application of random forest modeling for combined datasets of macromorphoscopic and craniometric data.

There were several limitations to this study. Primarily, since this was a matched dataset, samples that did not have both the macromorphoscopic observations and corresponding craniometric data were excluded from the sample. This depleted the overall sample size as all of the crania that was originally sampled had macromorphoscopic data, but did not have the craniometric data. However, it is important to highlight the strength of a matched data set. Because of this, there was no sampling bias, and individuals were not included to boost the number of samples for analysis of the macromorphoscopic data. Additionally, a few of the crania received a blank score for some of the macromorphoscopic traits due to poor preservation of the observed area for the specific trait. Although these individuals were included as one of the samples within the analysis, their incomplete nature was noted. The sample sizes were considerably smaller due to the elimination of individuals to create a matched data set.

The most difficult traits that may prove to have inconsistent scoring may include: inferior nasal aperture, nasal bone contour, and orbit shape. Generally, because nasal bone contour is instrumentally derived, there may be differences in scores due to the

placement of the contour gauge. For future research, interobserver error using the

macromacromorphoscopic data from Hefner (2009) and additional scoring from other

research studies should be compared to isolate which traits may be more troublesome to

assess and explore the reasoning behind these difficult traits.

V. CONCLUSION

Methods developed from cranial macromorphoscopic traits and craniometric variables inform forensic anthropologists how to proceed with building the biological profile for evaluating unidentified skeletal remains (Sauer and Wankmiller 2009; SWGANTH 2013). Within the literature, the macromorphoscopic and metric methods are especially useful for the estimation of ancestry. Most often qualitative analysis requires a complete visual assessment of the skeleton, normally the skull. Previously, macromorphoscopic traits have been typological in nature and have been assessed by simply describing features that are commonly observed in each population group and running through a trait list with specific traits that are ascribed to a particular population group (Christensen et al. 2014; Hefner 2009). These assessments can be deemed subjective and consensus between observers may be difficult. However, new steps are being introduced to strengthen this method by boosting this process with statistical analyses, an understanding of trait frequencies, and classification accuracies and error rates may contribute to the preexisting knowledge of variation between individuals (Hefner and Ousley 2014; Ousley and Hefner 2005). Metric methods have expanded farther than their macromorphoscopic counterparts. With the addition of reference groups for comparison, standardization of measurements, and ease of applying statistical analyses, this may be the preferred method for estimating ancestry.

New methods have been introduced to bring the macromorphoscopic methods to the level of their metric counterparts through powerful statistical analyses (Hefner and Ousley 2014; Hefner et al. 2014; Navega et al. 2013). A variety of machine learning methods were used to evaluate macromorphoscopc and craniometric data to evaluate the

classification accuracies of each technique and their general contribution to effectively estimating ancestry. In the present study, decision tree analysis, random forest modeling, artificial neural networks, support vector machines, and linear discriminant function analysis was applied to the macromorphoscopic, metric, and combined datasets. The high classification accuracies from the random forest model and the support vector analysis suggested that these machine-learning techniques are applicable to problems in forensic anthropology. The best machine learning methods for a combined analysis of qualitative and quantitative traits utilize support vector machines and random forest modeling. Further research must be conducted to assess the continued use of these analyses on combined datasets.

Macromorphoscopic traits can be analyzed with the statistical rigor that has been consistently seen in craniometric analysis (Hefner and Ousley 2014). Now both data types can be combined and analyzed alongside one another (Hefner et al 2014). Constant repetition and validation studies of these methods may be useful and can be applied to all aspects of the biological profile, which require classification and prediction methods (Christensen and Crowder 2009). Through this study, useful machine learning methods were identified and can be incorporated into a new toolkit for current and future forensic anthropologists. Much can be gleaned from discovering new statistical methodologies that arise through alternate sources, such as the origins of machine learning.

In the future, research involving ancestry estimation should be expanded to include samples from additional populations to generate trait frequencies for reference data and also for cross comparison. By compiling these new frequencies, a more comprehensive picture of the range of variation can be seen. Continued research using

machine-learning methods should be applied in future studies to analyze the various

classification accuracies and isolate particular machine-learning methods that work the

best for a particular type of dataset. Future studies may even incorporate postcranial data

to see if there are differences in classification accuracies between cranial and postcranial

measurements, especially for the purposes of ancestry estimation. The application of

machine learning techniques in forensic anthropology has made significant strides within

the past few years and will continue to shape the discipline as it continues to progress.

APPENDIX SECTION

APPENDIX A: DEFINITIONS OF MACROMORPHOSCOPIC TRAITS AND THEIR

CORRESPONDING CHARACTER STATES

Trait Cat Key	Trait Name	Character State Description
ANS	Anterior Nasal Spine	1 : <u>Slight </u>- minimal-to-no projection of the anterior nasal spine beyond the inferior nasal aperture. 2 : <u>Intermediate</u> – a moderate projection of the anterior nasal spine beyond the inferior nasal aperature 3: <u>Marked</u> – a pronounced projection fo the anterior nasal spine beyond the inferior nasal aperture.
INA	Inferior Nasal Aperture	1 : An inferior sloping of the nasal floor which begins within the nasal cavity and terminates on the vertical surface of the maxilla, producing a smooth transition. The morphology is distinct from INA 2 regarding the more posterior origin and the greater slope of INA 1. 2 : Sloping of the nasal aperture beginning more anteriorly than in INA 1, and with more angulation at the exit of the nasal opening. 3 : The transition from nasal floor to the vertical maxilla is not sloping, nor is there an intervening projection, or sill. Generally, this morphology is a right angle, although a more blunted form may be observed. 4 : Any superior incline of the anterior nasal floor, creating a weak (but present) vertical ridge of bone that traverses the inferior nasal border (partial nasal sill) 5 : A pronounced ride (nasal sill) obstructing the nasal floor-to-maxilla transition.
IOB	Interorbital Breadth	1 : A narrow IOB. 2 : A medium IOB. 3 : A broad IOB.

MT	Malar Tubercle	0 : No projection of bone
		1 : A trace tubercle below the ruler's edge (about 2 mm or less)
		2 : A medium protrusion below the ruler's edge (roughly 2 to 4 mm).
		3 : A pronounced tubercle below the ruler's edge (roughly 4 mm or more).
NAW	Nasal Aperture Width	1 : A narrow NAW.
		2 : A medium NAW.
		3 : A broad NAW.
NBC	Nasal Bone Contour	0 : Low and rounded nasal bone contour. NBC 1 presents a circular shape and lacks steep walls.
		1 : An oval contour, with elongated, high, and rounded lateral walls.
		2 : Steep lateral walls and a broad (roughly 7 mm or more), flat superior surface "plateau," noted on the contour gauge as a flat cluster of needles in the midline.
		3 : Steep-sided lateral walls and a narrow superior surface "plateau."
		4 : Triangular cross section, lacking a superior surface "plateau."
NBS	Nasal Bone Shape	1 : Nasal bones with no nasal pinch. The nasal bones may be wide or narrow.
		2 : Nasal bones with a superior pinch and minimal lateral bulging.
		3 : Nasal bones with superior pinch and pronounced lateral bulging of the inferior region.
		4 : Triangular-shaped nasal bones.
		**To differentiate between a score of 2 and 3, the amount of

		lateral bulging in the inferior region.
Novg	Nasal Overgrowth	0 : No overgrowth 1 : Any projection of the lateral border of the nasal bones (at nasale inferious) beyond the maxillary border.
OS	Orbital Shape	1 : <u>Rectangular</u> – Orbits with horizonal margins longer than vertical margins, but otherwise parallel. 2 : <u>Circular</u> – Orbital margin is approximately equidistant from center on all sides. 3 : <u>Rhombic</u> – Medial border height is shorter than lateral border height.
TPS	Transverse Palatine Suture	0 : If the right and left halves of the suture do not contact each other at midline, but the suture is otherwise straight, score the suture 0. 1 : The suture crosses the palate perpendicular to the median palatine suture, with no significant anterior or posterior deviation. 2 : The suture crosses the palate perpendicular to the medial palatine suture, but near this junction a significant anterior deviation, or bulging, is present. 3 : The suture crosses the palate, but deviates anteriorly and posteriorly (e.g., M-shaped) in the region of the median palatine suture. 4 : The suture crosses the palate perpendicular to the median palatine suture, but near this junction, a posterior deviation, or bulging, is present.

B: DEFINITION OF PALATE SHAPE AND ITS THREE STATES

Trait Cat Key	Trait Name	Character State Description
PS	Palate Shape	1. Elliptic - Smooth, round curvature of the anterior portion of the palate combined with a mid-arch (ca. M1 and M2) widening relative to M3, contributing to the appearance of constricted (medially-positioned) 3rd molars. The curvature of the dental arcade is even throughout, and the 3rd molars have a "tucked-in" appearance. If projected beyond the M3s, the left and right mesio-distal midlines would converge posterior to the palate. This morphology closely resembles an incomplete ellipse.

2. Parabolic – Smooth, rounded curvature of the anterior portion of the palate, combined with an even, gradual flaring of the posterior dentition. The curvature of the dental arcade is most noticeable anteriorly, with a gradual, but not total, straightening of the dental arch curvature posteriorly. If projected beyond the M3s, the left and right mesio-distal midlines would continue to diverge posteriorly. This morphology closely resembles a true geometric parabola. This form displays the greatest variation in width relative to palate length; however, the continued divergence of the posterior curvature is the critical feature in identifying this character state.

3. Hyperbolic – Smooth, slightly flattened curvature of the anterior portion of the palate, combined with a straight, more-or-less parallel configuration of the posterior portions of the arch. Curvature is most marked in the vicinity of the canines, and becomes nearly straight through the posterior dentition. If projected beyond the M3s, the left and right mesio-distal midlines may diverge slightly, but to a lesser degree than in the "parabolic" form. This morphology closely resembles an inverted U-shape. |

APPENDIX C: GRAPHIC OF PALATE SHAPE MORPHOLOGY
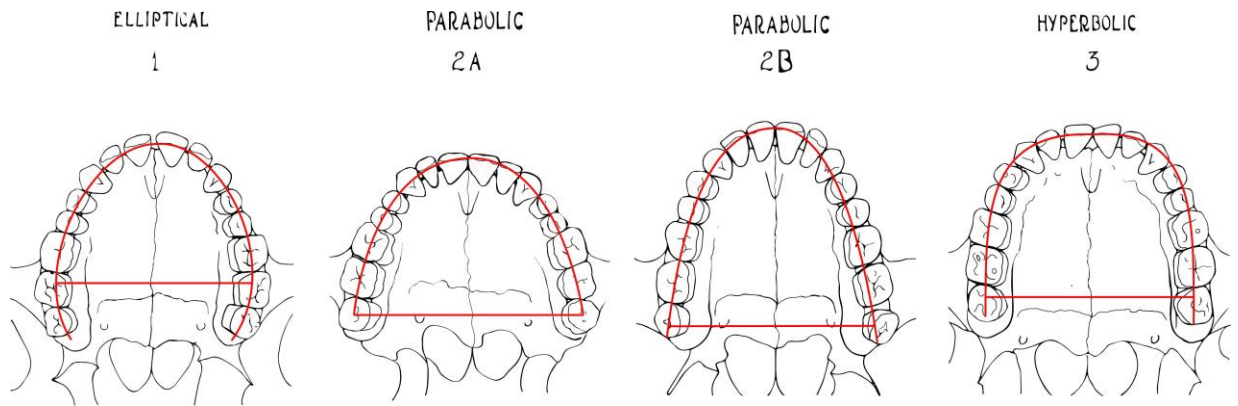
**Image courtesy of Rachel Canfield

APPENDIX D: DEFINITIONS OF CRANIOMETRIC MEASUREMENTS USED

	Craniometric Measurement Code	Name	Description of Measurement
1	GOL	Glabello-occipital length	Greatest length from the glabellar region, in the median sagittal plane.
2	NOL	Nasio-occipital length	Greatest cranial length in the median sagittal plane, measured from nasion.
3	BNL	Cranial based length	Direct length between nasion and basion.
4	BBH	Basion-bregma height	Distance from bregma to basion, as defined.
5	XCB	Maximum cranial breadth	The maximum cranial breadth perpendicular to the median sagittal plane (above supramastoid crests).
6	XFB	Maximum frontal breadth	The maximum breadth at the coronal suture, perpendicular to the median plane.
7	WFB	Minimum frontal breadth	The direct distance between the two frontotemporale.
8	ZYB	Bizygomatic diameter	The maximum breadth across the zygomatic arches, wherever found, perpendicular to the median plane.
9	AUB	Biauricular breadth	The least exterior breadth across the roots of the zygomatic processes, wherever found.
10	ASB	Biasterionic breadth	Direct measurement from one asterion to the other.
11	BPL	Basion-prosthion length	The facial length from prosthion to basion, as defined.
12	NPH	Nasion-prosthion length	Upper facial height from nasion to prosthion, as defined.
13	NLH	Nasal height	The average height from nasion to the lowest point on the border of the nasal aperture on either side.
14	JUB	Bijugal breadth	The external breadth across the malar at the juglia , i.e., at the deepest points in the curvature between the frontal and temporal process of the malars.
15	NLB	Nasal breadth	The distance between the anterior edges of the nasal aperture at its widest extent.
16	MAB	Palate breadth	The greatest breadth across the alveolar borders, wherever found, perpendicular to the median plane.
17	MAL	Maxillo-aleveolar length	Direct distance from prosthion to alveolon

18	MDH	Mastoid length	The length of the mastoid process below, and perpendicular to the eye-ear plane in the vertical plane.
19	OBH	Orbital height	The height between the upper and lower borders of the left orbit, perpendicular to the long axis of the orbit and bisecting it.
20	OBB	Orbital breadth	Breadth from ectoconchion to dacryon, as defined, approximating the longitudinal axis, which bisects the orbit into equal upper and lower parts.
21	DKB	Interorbital breadth	The breadth across the nasal space from dacryon to dacryon.
22	NDS	Naso-dacryal substense	
23	WNB	Simiotic chord (least nasal breadth)	
24	SIS	Simiotic subtense	
25	ZMB	Bimaxillary breadth	
26	SSS	Zygomaxillary subtense	
27	FMB	Bifrontal breadth	
28	NAS	Nasio-frontal subtense	
29	EKB	Biorbital breadth	The breadth across the orbits from ectoconchion to ectoconchion.
30	DKS	Dacryon subtense	
31	IML	Malar length, inferior	
32	XML	Malar length, maximum	
33	MLS	Malar subtense	
34	WMH	Cheek height	
35	GLS	Glabella projection	
36	STB	Bistephanic breadth	
37	STS		
38	FRC	Frontal chord	The frontal chord, or direct distance from nasion to bregma, taken in the midplane and at the external surface.
39	FRS	Nasion-bregma subtense (Frontal subtense)	
40	FRF	Nasion-subtense fraction	
41	PAC	Parietal chord	The external chord, or direct distance from bregma to lambda taken in the midplane and at the external surface.

42	PAS	Bregma-lambda subtense (Parietal subtense)	
43	PAF	Bregma-subtense fraction	
44	OCC	Occipital chord	The external occipital chord, or direct distance from lambda to opisthion taken in the midplane and at the external surface.
45	OCS		
46	OCF		
47	FOL	Foramen magnum length	The length from basion to opisthion, as defined.
48	FOB	Foramen magnum breadth	The distance between the lateral margins of foramen magnum at the points of greatest lateral curvature.
49	NAR	Nasion radius	
50	SSR	Subspinale radius	
51	PRR	Prosthion radius	
52	DKR	Dacryon radius	
53	ZOR	Zygoorbitale radius	
54	FMR	Frontomalare radius	
55	EKR	Ectoconchion radius	
56	ZMR	Zygomaxillare radius	
57	AVR	Molar 1 Aleveolus radius	
58	BRR	Bregma radius	
59	VRR	Vertex radius	
60	LAR	Lambda radius	
61	OSR	Opisthion radius	
62	BAR	Basion radius	
63	NAA		
64	PRA		
65	BAA		
66	NBA		
67	BBA		
68	BRA		
69	SSA		
70	NFA		
71	DKA		
72	NDA		
73	SIA		
74	FRA		
75	PAA		

76	OCA		
77	RFA		
78	RPA		
79	ROA		
80	BSA		
81	SBA		
82	SLA		
83	TBA		
84	UFHT	Upper facial height	The direct distance from nasion to prosthion.
85	UFBR	Upper facial breadth	

REFERENCES CITED

Anderson BE
  2008 Identifying the Dead: Methods Utilized by the Pima County (Arizona) Office of
      the Medical Examiner for Undocumented Border Crossers: 2001-2006. Journal
      of Forensic Sciences 53(1):8-15.

Anderson BE
  1998 "Forensic Anthropology As a Science: Is There a Difference Between Academic
      and Applied Uses of Biological Anthropology?" Doctoral Dissertation. The
      University of Arizona.
      http://arizona.openrepository.com/arizona/handle/10150/282649.

Anderson BE and Parks BO.
  2008 Symposium on Border Crossing Deaths: Introduction. Journal of Forensic
      Sciences 53(1):6-7.

Birkby WH, Fenton TW and Anderson BE
  2008 Identifying Southwest Hispanics Using Nonmetric Traits and the Cultural Profile.
      American Journal of Physical Anthropology 53(1):29-33.

Breiman L
  2004 Consistency For A Simple Model Of Random Forests. Technical Report 670.
      University of California at Berkeley.

Brues, AM.
  1990 The Once and Future Diagnosis of Race. Skeletal Attribution of Race.
      Albuquerque, NM: University of New Mexico: Maxwell Museum of
      Anthropological Papers No. 4. p 1-9.

Buikstra JE and Ubelaker DH, editors.
  1994 Standards for Data Collection from Human Skeletal Remains: Proceedings of a
      Seminar at the Field Museum of Natural History. Fayetteville: Arkansas
      Archaeological Series.

Burris BB and Harris EF
  1998 Identification of Race and Sex from Palate Dimensions. Journal of Forensic
      Sciences 43(5):959-963.

Byers SN, Churchill SE and Curran B
  1997 Identification of Euro-Americans, Afro-Americans, and Amerindians from Palatal
      Dimensions. Journal of Forensic Sciences 42(1):3-9.

Carpenter JC
  1976 A Comparative Study of Metric and Nonmetric Traits in a Series of Modern
      Crania. American Journal of Physical Anthropology 45:337-343.

Chapman P
  1991 The palate: Race and sex. Unpublished manuscript on file at the University of
      Wyoming.

Chapman P and Gill GW
  1993 Use of the palate and palatine suture in race identification. Paper presented at the
      45th Annual Meeting of the Academy of Forensic Sciences.

Cheverud JM, Buikstra JE and Twichell E
  1979 Relationships between Non-Metric Skeletal Traits and Cranial Size and Shape.
      American Journal of Physical Anthropology 50:191-198.

Christensen AM and Crowder CM
  2009 Evidentiary Standards for Forensic Anthropology. Journal of Forensic Sciences
      54(6):1211-1216.

Christensen AM, Passalacqua NV and Bartelink EJ
  2014 Ancestry Estimation. *In* Forensic Anthropology: Current Methods and Practice.
      Pp. 223-242. San Diego: Academic Press.

Corruccini RS
  1976 The Interaction Between Nonmetric and Metric Cranial Variation. American
      Journal of Physical Anthropology 44:285-293.

Corruccini RS
  1974 An Examination of the Meaning of Cranial Discrete Traits for Human Skeletal
      Biological Studies. American Journal of Physical Anthropology 40:425-445.

Corsini MM, Schmitt A, Bruzek J
   2005 Aging process variability on the human skeleton: artificial network as an
      appropriate tool for age at death assessment. Forensic Sci Int 148:163–167.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

DiGangi, EA and Hefner JT
  2013 Ancestry Estimation. *In* Research Methods in Human Skeletal Biology.
      Elizabeth A. Digangi and Megan K. Moore, eds. Pp. 117-150. Waltham:
      Academic Press.

D'Orazio M
  2014 Statistical Matching and Imputation of Survey Data with StatMatch. Italian
      National Institute of Statistics. ESSnet Project. Funded by Eurostat.

Du Jardin P, Ponsaillé J, Alunni-Perret V, Quatrehomme G.
  2014 A comparison between neural network and other metric methods to determine
      sex from the upper femur in a modern French population. Forensic Science
      International 192(127):e1–e6.

 Ferguson E, Kerr N and Rynn C
  2011 Race and Ancestry. *In* Forensic Anthropology 2000 to 2010. Sue Black and
      Eilidh Ferguson, eds. Pp. 119-153.   Boca Raton: Taylor & Francis Group.

Fleiss JL
  1971 Measuring nominal scale agreement among many raters Psychological
      Bulletin.76(5):378–382.

Forensic Anthropology Center at Texas State (FACTS)
      2014 Texas State University http://www.txstate.edu/anthropology/facts/ (accessed
      October 1, 2014).

Galdames ICS, Matalama DAZ, and Smith RL
  2008 Accuracy of Palate Shape as Sex Indicator in Human Skull with Maxillary Teeth
      Loss. International Journal of Morphology 26(4):989-993.

Giles, E. and Elliot O.
  1962 Race Identification from Cranial Measurements. Journal of Forensic Sciences
      7:147-157.

Gill GW
  1998 Craniofacial Criteria in the Skeletal Attribution of Race. *In* Forensic Osteology:
      Advances in the Identification of Human Remains. Kathy Reichs, ed. Pp. 239-
      315. Springfield: Charles C. Thomas Publisher.

Gilbert BM and Gill GW
  1990 A Metric Technique for Identifying American Indian Femora. In: Gill G, and
      Rhine S, editors. Skeletal Attribution of Race. Albuquerque, NM: University of
      New Mexico: Maxwell Museum of Anthropological Papers No. 4. p. 97-99.

Hefner JT
   2012 Macromorphoscophics. In: Wilczak, Cynthia A. and Dudar, J. Christopher,
      editors. Osteoware Software Manual Volume I. Washington, D.C.: Smithsonian
      Institution.

Hefner JT
  2011 Cranial non-metrics and macromorphoscopics in OsteoWare. American
      Academy of Physical Anthropology. 144:159.

Hefner JT
  2007 "The Statistical Determination of Ancestry Using Cranial Nonmetric Traits."
    Doctoral Dissertation, University of Florida,
    http://ufdc.ufl.edu/UFE0021200/00001.

Hefner JT
  2009 Cranial Nonmetric Variation and Estimating Ancestry. Journal of Forensic
    Sciences 54:985-995.

Hefner JT and Ousley SD
  2014 Statistical Classification Methods for Estimating Ancestry Using Morphoscopic
    Traits. Journal of Forensic Sciences 1-8.

Hefner JT and Ousley SD
  2006 Morphoscopic traits and the statistical determination of ancestry II. *57th Annual*
    Proceedings of the American Academy of Forensic Sciences 21-26.

Hefner JT, Ousley SD, and Dirkmaat DC
  2012 Morphoscopic Traits and the Assessment of Ancestry. *In:* A Companion to
    Forensic Anthropology. Dennis C. Dirkmaat, ed. Pp: 287-310. Sussex: Wiley-
    Blackwell.

Hefner JT, Spradley MK and Anderson BE
  2014 Ancestry Assessment Using Random Forest Modeling, Journal of Forensic
    Sciences 1-7.

Hinkes MJ
  1993 Race, Ethnicity, and Forensic Anthropology: Realities of Racial Determination in
    a Forensic Setting National Association for the Practice of Anthropology (NAPA)
    Bulletin 13:48-53.

Hooton, EA
  1926 Lecture notes of EA Hooton. Peabody Museum archives. Cambridge, MA.

Howells WW
  1973 Cranial Variation in Man – A Study by Multivariate Analysis of Patterns of
    Difference Among Recent Human Populations. Cambridge: Harvard University
    Press.

Howells WW
  1989 Skull Shapes and the Map. Craniometric Analyses in the Dispersion of Modern
    Homo. Papers of the Peabody Museum of Archaeology and Ethnology, vol. 79,
    pp. 189. Cambridge, Mass.: Peabody Museum.

Hughes CE, Juarez CA, Hughes TL, Galloway A, Fowler G, Chacon S
  2011 A simulation for exploring the effects of the "trait list" method's subjectivity on consistency and accuracy of ancestry estimations.   Journal of Forensic Sciences 56(5):1094-1106.

Hurst CV
  2012 Morphoscopic Trait Expressions Used to Identify Southwest Hispanics. Journal of Forensic Sciences 57(4):859-865.

Hunt DR and Albanese J
  2005 History and demographic composition of the Robert J. Terry anatomical collection. American Journal of Physical Anthropology 127(4): 406-417.

Hunt DR and Spatola B
  2008 History and demographic profile of the George S. Huntington collection at the Smithsonian Institution, American Journal of Physical Anthropology 121-122.

Jantz RL and Moore-Jansen PH
  1984 Forensic Data Bank. University of Tennessee, Knoxville, TN.

Jantz RL and Ousley SD
  2005 FORDISC 3: Computerized Forensic Discriminant Functions. University of Tennessee, Knoxville, TN.

Klales AR and Kenyhercz MW
  2014 Morphological Assessment of Ancestry using Cranial Macromorphoscopics. Journal of Forensic Sciences doi: 10.1111/1556-4029.12563.

Komar DA and Grivas C
   2008 Manufactured populations: What do contemporary reference collections represent? A comparative study using the Maxwell Museum documented collection. American Journal of Physical Anthropology 137(2):224-233.

Klepinger LL
  2006 Fundamentals of Forensic Anthropology. John Wiley & Sons Inc.: Hoboken.

Krogman, WM and Işcan MY
  1986 The Human Skeleton in Forensic Medicine. Charles C. Thomas: Springfield.

L'Abbé EN, Van Rooyen C, Nawrocki SP and Becker PJ
  2009 An evaluation of non-metric cranial traits used to estimate ancestry in a South African sample. Forensic Science International 209(1-3).

Mahfouz M, Badawi A, Merkl B, Fatah EE, Pritchard E, Kesler K, Moore M, Jantz R, Jantz L.
   2007 Patella sex determination by 3D statistical shape models and non-linear classifiers. Forensic Science International. 173:161-170.

Mavroudas SR, Wescott DJ, Spradley MK, Hamilton MD, and Stull KE
  2012 The Forensic Anthropology Center at Texas State University. *64th Annual Proceedings of the American Academy of Forensic Sciences* 377-378.

McBride DG, Dietz MJ, Vennemeyer MT, Meadors SA, Benfer RA, Furbee NL.
  2001 Bootstrap Methods for Sex Determination from the Os Coxae Using the ID3 Algorithm. Journal of Forensic Sciences 46(3):427-431.

Maier CA
  2013 "Palate Shape and Depth: A Shape Matching and Machine Learning Method for Assessment of Ancestry from Skeletal Remains." Master's Thesis, Louisiana State University, http://etd.lsu.edu/docs/available/etd-03182013 190045/unrestricted/maierwholethesis.pdf.

Napoli ML and Birkby WH
  1990 Racial differences in the visibility of the oval window in the middle ear. In: Gill G, and Rhine S, editors. Skeletal Attribution of Race. Albuquerque, NM: University of New Mexico: Maxwell Museum of Anthropological Papers No. 4.

Navega D, Coelho C, Vicente R, Ferreira MT, Wasterlain S, and Cunha E
  2013 AncesTrees: ancestry estimation with randomized decision trees. *International journal of legal medicine* (2013): 1-9.

Navega D., Vicente R., Viera DN., Ross AH, and E. Cunha.
  2014. Sex estimation from the tarsal bones in Portuguese sample: a machine learning approach. International Journal of Legal Medicine. 1-9.

Osteoware [computer program]
  2011 Standardized skeletal documentation software. Washington, DC: Smithsonian Institution   National Museum of Natural History, http://osteoware.si.edu/ (accessed July 8 2013).

Ousley SD, Jantz, RL, and Freid D.
  2009 Understanding Race and Human Variation: Why Forensic Anthropologists are Good at Identifying Race, American Journal of Physical Anthropology 139:68-76.

Ousley SD and Hefner JT
  2005 The statistical determination of ancestry. Proceedings of the 57th Annual Meeting of the American Academy of Forensic Sciences; 2005 Feb 21–26; New Orleans, LA. Colorado Springs, CO: American Academy of Forensic Sciences.

Rafter, N
   2004 Earnest A. Hooton and the Biological Tradition in American Criminology,
       Criminology   42(3):735-772.

Relethford J
   2009 Race and Global Patterns of Phenotypic Variation, American Journal of Physical
       Anthropology 139:16-22.

Rhine SJ
   1993 Skeletal Criteria for Racial Attribution. National Association for the Practice of
       Anthropology (NAPA) Bulletin 13:54-67.

Rhine SJ
   1990 Nonmetric Skull Racing. In: Gill G, and Rhine S, editors. Skeletal Attribution of
       Race.   Albuquerque, NM: University of New Mexico: Maxwell Museum of
       Anthropological Papers No. 4.

Richtsmeier JT, Cheverud JM, and Buikstra JE.
   1984 The Relationship Between Cranial Metric and Nonmetric Traits in the Rhesus
       Macaques From Cayo Santiago. *American Journal of Physical Anthropology*
       64:213-222.

Sauer NJ
    1992 Forensic Anthropology and the Concept of Race: If Races Don't Exist, Why
       Are Forensic Anthropologists So Good at Identifying Them, Social Science &
       Medicine 34:107-111.

Sauer NJ and Wankmiller JC
   2009 The Assessment of Ancestry and the Concept of Race. *In* Handbook of Forensic
       Anthropology and Archaeology. Soren Blau and Douglas Ubelaker, eds. Pp.
       187-200. Walnut Creek: Left Coast Press.

Scientific Working Group for Forensic Anthropology (SWGANTH)
   2013 Ancestry Assessment.

Shirley NR, Wilson RJ, and Jantz LM
   2011 Cadaver use at the University of Tennessee's Anthropological Research Facility.
       Clinical Anatomy 24(3):372-380.

Spradley MK
   2013 Project IDENTIFICATION: Developing Accurate Identification Criteria for
       Hispanics. NIJ Grant: 2008-DN-BX-K464.
       https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=266273 (accessed
       September 24, 2014).

Spradley, MK, Jantz, RL, Robinson A, and Peccerelli F
  2008 Demographic Changes and Forensic Identification: Problems in Metric
    Identification of Hispanic Skeletons. Journal of Forensic Sciences 53(1):21-28.

SPSS, Inc.
  2013 SPSS 22.0. Chicago, Illinois.

StatSoft
  2014 Statistica Electronic Manual.
    http://documentation.statsoft.com/STATISTICAHelp.aspx?path=common/About
    STATISTICA/ElectronicManualIndex (accessed October 15, 2014).

Tise, ML, Spradley, MK, and Anderson BE
  2013 Postcranial Sex Estimation of Individuals Considered Hispanic, Journal of
    Forensic Sciences 58(S1):S9-S14.

University of New Mexico (UNM)
  n.d. Laboratory of Human Osteology, Maxwell Museum of Anthropology: Documented
    Skeletal Collection. http://www.unm.edu/~osteolab/coll_doc.html (accessed June
    30, 2013).

Vitek CL
  2012 "A Critical Analysis of the Use of Non-Metric Traits for Ancestry Estimation
    Among Two North American Population Samples. " Master's Thesis, University
    of Tennessee, http://trace.tennessee.edu/utk_gradthes/1218.

Wescott DJ
  2005 Population variation in femur subtrochanteric shape Journal of Forensic
    Sciences 50:286–293.

Wheat A
  2009 "Assessing Ancestry Through Non-Metric Traits of the Skull: A Test of
    Education and Experience." Master's Thesis, Texas State University,
    https://digital.library.txstate.edu/handle/10877/4052.

Williams BA and Rogers TL
  2006 Evaluating the Accuracy and Precision of Cranial Morphological Traits for Sex
    Determination, Journal of Forensic Sciences 51(4):729-735.

Williams G
  2011 Data Mining with Rattle and R: The Art of Excavating Data for Knowledge
    Discovery Springer Science: New York.

Wilson C
  2010 The Relationship between Size and Expression of Nonmetric Traits on the Human
    Skull, The Arbutus Review 1:81-97.