June 2021

The Science of Guessing: Critiquing Ancestral Estimation Through Computer Generated Statistical Analysis Within Forensic Anthropology in a Real-World Setting

Christopher J. Turner
University of South Florida

The Science of Guessing: Critiquing Ancestral Estimation Through Computer Generated Statistical Analysis Within Forensic Anthropology in a Real-World Setting

by

Christopher J. Turner

A thesis submitted in partial fulfillment
of the requirements for the degree of
Masters of Arts in Applied Anthropology
with a concentration in Biological Anthropology
Department of Anthropology
College of Arts and Sciences
University of South Florida

Major Professor: Erin H. Kimmerle, Ph.D.
Christian E. Wells, Ph.D
Jonathan Bethard, Ph.D

Date of Approval
May 27, 2021

Keywords: Craniometrics, Fordisc 3.1, 3D-ID, Population Affiliation, Geographic Ancestry

ACKNOWLEDGMENTS

This article has been part of my master's for Applied Anthropology with a Forensic Anthropology focus. This study would not have been possible without the assistance of my committee, Dr. Kimmerle, Dr. Wells, and Dr. Bethard; their insight has been critical. I also wish to thank the other graduate students in our lab who took the time to ensure my competence with Fordisc 3.1 and 3D-ID, Kelsee Hentschel-Fey, Liotta Noche-Dowdy, and Gennifer Goad. I would also like to acknowledge the support and guidance of Jacquelyn Heuer and Fiona Felker. They have continued to prepare me for the pitfalls of graduate school and have been invaluable counsel throughout. Finally, I wish to thank my wife, Rya, and two sons, Ronan and Stirling, who have given me focus on the task at hand.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Research on current methods of ancestral estimation must reflect on biological heritage to aid in human identification. Using modern craniometrics methods, how do individuals with a varied biological history affect ancestral estimation? Today, the most used and reliable methods for craniometrics analysis for ancestral estimation in forensic anthropology are computer programs. Two programs are analyzed in this study, Fordisc 3.0 and 3D-ID. The analysis of these computer programs goes beyond the controlled environment provided by an osteological collection. These remains of individuals were unidentified, only to be identified later, through academic research, police work, and public outreach. The selection of samples occurred if they fit the following two criteria: 1) To evaluate known ancestry, the victim was initially unidentified but was later positively identified, and 2) both Linear and 3D coordinate craniometric data was available to test Fordisc 3.1 and 3D-ID. Humans adapt to their environment biologically and culturally, identifying with familiar cultures, foods, objects, events and, how we look. Thus, ancestral components to a person's appearance can help outline the parameters in a search to return a lost loved one to their family and finish the last chapter in an individual's life. We establish any trends in the correct and incorrect estimations by analyzing the posterior probability (pp) and typicalities (typ). Both computer programs struggled with the "Hispanic" cohort placement while finding higher reliability in European Americans' estimations than any other ancestral group for both 3D-ID and Fordisc 3.1.

CHAPTER ONE:
INTRODUCTION

A critical component of the biological profile is classifying unidentified remains into ancestral categories, which is often considered the most complicated element (1). Practitioners commonly take a broad "ancestral" approach. Geographical regions are the foundation for these estimations, such as Africa, Asia, and Europe; this has little basis on evolutionary factors and no genuine population affinity (2). Consequently, there has been a push within forensic anthropology to take an evolutionary approach, defined as population affiliation (4,5). Population affiliation is ultimately geographical ancestry (3), connecting ancestral history to geographical regions, acknowledging both evolutionary pressures and forces. These factors create human variation based on socio-cultural influences, allele frequencies distributed by clines, and migratory pressures, all accumulating to affiliate a population through time to a geographical location.

Consequently, ancestry estimations should focus on traits with known heritability and evolutionary (environmental) relationships, which results in a more precise population designation. This precision gives a more accurate indication of biodistance, which is the deviation of groups relatedness through geography and time (6). Numerous approaches for estimating ancestry include linear and three-dimensional analyses of craniometric variables and analyses of morphoscopic traits and dental morphology (7–12). Today, it is well established that biological "race" does not exist (13,14), although race (and ancestry) remain as cultural

constructs and contribute to the identification of unknown persons. Culturally, people identify with their ancestry and heritage in various ways; other cultures and one's genetic history may contradict those self-beliefs.

Moreover, interactions between heredity and one's environment significantly influence factors surrounding a person's appearance: from eye and skin color to the shape of facial features, hair texture, height, etcetera. More recently, research highlights differences in individuals' facial morphology through single nucleotide polymorphisms (SNPs) (15). Changes in DNA through mutation, adaptation, and mate selection demonstrate how populations evolve. The forces of evolution utilize the biological building blocks available to them, pressured and constrained by the local environment. Research on current methods used to estimate population affiliation to reflect biological heritage accurately is critical for human identification (16). Evaluating the biodistance of populations and the classification of an unknown person into the most closely related group(s) is well established (7-8). However, how can accuracy be improved by altering the reference populations or the specific measurements used?

Currently, the most commonly used and reliable methods for estimating population affiliation through craniometrics are two computer programs (17–19), Fordisc 3.1 (20), and 3D-ID (21). As demonstrated in numerous studies, these programs have high success rates, indicating 70-90% accuracy (18,22–25). Both programs use craniometric data to establish best fit grouping with canonical variant analysis. 3D-ID utilizes three-dimensional coordinate data through geometric morphometrics (GM) to estimate population affiliation by relating the measurements to stored data within the program, outlining a shared population history via biodistance. In contrast, Fordisc 3.1 utilizes linear craniometrics. Additionally, both programs utilize Mahalanobis distance (D^2), discriminant function analysis (DFA), and Linear

Discriminant Function (LDF) (26). DFA is used to establish where an individual falls within the centroids (the mean score of the canonical), which is a measure of best fit and indicates how strongly aligned the unknown individual is to the population.

Fordisc 3.1 uses measurements from the Forensic Data Bank (FDB), beginning in 1986 by a grant from the National Institute of Justice (NIJ) (20). The FDB relies on eight populations, with all individuals born after 1930. The reference populations, as defined by Fordisc 3.1, include: American Blacks (n=224 males, n=137 females); American Indian (n=59 males, n=32 females); American Whites (n=737 males, n=454 females); Chinese (n=80 males, n=0 females); Guatemalan (n=83 males, n=0 females); Hispanics (n=281 males, n=74 females); Japanese (n=84 males, n=58 females); and Vietnamese (n=51 males, n=0 females).

The FDB measurements are exploited in DFA statistical procedures to determine LDF. This program allows the practitioner to analyze multiple ancestral groups and estimates where the combined dimensions best fit into the canonical variants. The canonical variates analysis compares the measurements to multiple centroids. These centroids are created and defined by the mean scores of the reference measurements used for each ancestral group selected.

3D-ID utilizes a wide range of population data from numerous forensic laboratories and museum collections throughout the world. There are also requirements for known demographic data for inclusion into the database, such as ancestry, age, and sex. Currently, the database consists of n=2,508 individuals (Table 1). The critical differences between the two listed programs are different reference populations, the terminology of population groups, and the fact that 3D-ID uses both shape and size of the skull to estimate relatedness, rather than just size. It is important to note that the creators of 3D-ID found common ancestral terms to have little

biological meaning and thus chose group identification related to geographical regions to acknowledge clinal variation, utilizing the categories listed in Table 1.

Both programs are similar in their designs in that the practitioner must select which populations and sex groups to include in the analysis. It is at the practitioner's discretion to choose populations within the program, thereby setting the parameters to compare and identify the best fit between an unknown individual and population. However, as Jantz and Ousley point out (20), these statistical analyses will be less effective with more reference populations utilized within any research. Therefore, results will vary based on the populations selected, and it is up to the practitioner to use the program to choose the "best" populations for comparison. If there are too many reference populations, the practitioner runs the risk of a muddled analysis; too few, the practitioner runs the risk of excluding the affiliated population by removing it.

The purpose of this study is to investigate how population selection influences the results of ancestry estimation using the two craniometric methods discussed. The study is not a direct comparison of 3D-ID and Fordisc 3.1, but an analysis of how these computer programs are mailable, focusing on individuals with known "mixed ancestry." The individual cases are studied to understand how better practitioners can interpret population affinity. Here, the concept "mixed ancestry" refers to individuals whose parents came from two different reference populations.

Table 1. List of 3D-ID reference populations and their origin (27).

Population	Total size (n)	Source of Collection
African	27	Ross. Morton Collection, Smithsonian National Museum of Natural History (NMNH)
African American	272	Terry Collection, Pound Lab, Kimmerle, Ross, Maxwell Museum
African Brazilian	55	Urbanova
Brazilian	125	Urbanova
Circumcaribbean	26	Ross
Colombian	71	Bethard, DiGangi, Medellin
East African	36	Ross, American Museum of Natural History (AMNH)
East Asian	28	Berg, Ross, Morton Collection
European American	378	Ross, Kimmerle, Westcott, Terry Collection, Forensic Cases, Maxwell Museum
European Central	412	Urbanova
European Eastern	2	Urbanova
European Southeastern	266	Ross, Urbanova
European Southwestern	446	Ross
Japanese Brazilian	27	Urbanova
Mesoamerican	89	Anderson, Spradley, Ross
Nigeria	30	Ross, AMNH
South American	82	Ross
Syrian	43	Ross, AMNH
West African	93	Ross, AMNH
Total	**2508**	

CHAPTER TWO:

LITREATURE REVIEW

Currently, the field of forensic anthropology categorizes the grouping of individuals over time as ancestry. The verbiage to describe in-group and out-group classifications surrounding ancestry has not always been classified as such. Forensic anthropology is heavily influenced by how the legal system interprets an individual's categorized ancestry. This categorization was done initially as a scientific definition based on taxonomic classifications. Though forensic anthropology may no longer be in its infancy, it is still young compared to its parent fields of physical anthropology and human anatomy. As such, we draw much of our methods and understanding from anatomists and medical doctors throughout history. The father of American Forensic Anthropology, Thomas Dwight (1843-1911) from Harvard, identified skeletal remains for legal matters in the late 1800s (28). Dwight focused on sex, age, and stature and less on the concept of ancestry or race. The concept of race is not new and has seen much debate since Carl Linnaeus (1707-1778) created a taxonomic system defining species of plants and animals, including human classification.

Linnaeus grouped individuals into four categories: American, European, Asian, and African. His categorical divisions were based on five components: geographical location, skin color, humor (blood, bile, and phlegm), posture, and custom (14). Thus, the concept of race was given a firm grounding. The concept was founded on *The Great Chain of Being*, defining God at

the pinnacle and the lowest lifeforms at the bottom. This fundamental ideal maintains in today's western cultures (29).

Ethnocentric and colonial ideas were soon entrenched into the ideas of race. Johann Friedrich Blumenbach (1752-1840), who is considered the Father of Physical Anthropology, exemplified these colonial overtones (14). He classified humans into five groupings: Caucasian, Mongolian, Ethiopian, American, and Malay. Caucasian was based on his proposed origins for Europeans, a region known as the Caucus Mountains. This area Blumenbach described as the most beautiful region he had ever witnessed. Thus, it had to be the origin of the European race. Blumenbach did not shy from the idea that this was a categorical system based on a hierarchy with Caucasians at the pinnacle.

Races were considered fixed with no secular change, something Charles Darwin had refuted in the book *On the Origin of Species* (1859). Frans Boaz had disproven by studying the changes in immigrant's cranial measurements over time (30). The concept of race saw three distinct focal areas: essentialism, cladistic thinking, and biological determinism (31). Essentialism defined race, highlighting components that appeared to have scientific merit. Clades fueled the Nazi regime and the eugenics program defining some races as better than other races. Biological determination saw racial differences as evolutionary, and this evolution explained cultural differences. Thus, certain groups could evolve at faster rates and become civilized as opposed to savages (32).

It is no surprise that Ashley Montagu, in the early 1960s, identified the idea of race as a charged word, triggering a reaction from the general populous of western cultures (33). Like many American anthropologists, Montagu had witnessed the dangers of racism through eugenics in the States and the Nazi Party in Germany. In the 1960s, a further racist threat fixated the field,

as the civil rights movement gained momentum against segregation. Many stood up and denounced the concept of race; as early as 1934, the then President of the AAA, J.B.S. Haldane would speak out about the dangers of racism almost as a foreshadow of the tyranny the Nazi Germany Party would bring to the world some five years later (31). The shift saw a rise in the field of physical anthropology, vocally denouncing the concept of race. Thus, the concept of environmental pressures attributed to populations' evolution gained firm ground as these populations adapted to the environment around them. Charles Darwin, in 1859, had outlined the pressures of natural selection, citing species need to adapt to their environment to survive (34).

Frank Livingstone (1928-2005) and Sherwood Washburn saw that variation within the human population is based on genetic variation, significantly different within a local population (35). Livingstone promoted the idea of clines and defined them as follows:

1. *the recent advance of an advantageous gene*
2. *gene flow between populations which inhabit environments with different equilibrium frequencies for the gene*
3. *a gradual change in the equilibrium value of the gene along the cline. The theoretical analysis of clines has barely begun but there seems to be no need for the concept of race in this analysis.* (35)

To understand clinal variation, one needs to look at skin color, as this is the most visually identifiable example of clinal variation. Closer to the poles, skin variation is lighter to allow the synthesis of vitamin B12; on the equator, skin tone is darker due to the need for protective pigmentation from the sun's harsh rays (36). Montagu echoed these sentiments, and he outlines the forces of change: mutation, natural selection, genetic drift, and isolation. These allow humans to continue to change. However, there was high mobility in humans with social selection

pressures, ensuring continued migration and admixture to static races. Thus, there was little to support the concept of race in the eyes of these individuals.

There is still the idea that we do identify those around us based on the extremes of the clines we see. The question then raised is: if there is a gradual variant, why has the definition of race been so prominent? Brace argues a traveler would see a wide variance in groups due to the maritime trade's advent in the 12th and 13th centuries (13). Setting sail from Northern Europe and arriving in Africa or the Caribbean would show how different we could be. Thus, this has been a difficult concept to break, especially with the general public. Subsequently, there have been four distinct ideas regarding the concept of race, especially in forensic anthropology (37). The first belief considered race to be a natural category, or more importantly, treated it as such. Stanley Rhine asserts that secular change is a slow process and chastised any forensic anthropologist who would shy away from the use of race:

> *One should at least wistfully entertain the hope that society at the cusp of the 21st century would have progressed beyond the petty aspersions of racism to the recognition that, as with any species, our survival depends upon diversity* (38)

However, this philosophy of race being a natural category has been dismissed even by Rhine. Thus, a more popular theory identifies race as a flawed concept with a small error. While apparent on a larger scale, this error can be helpful as a tool in the same way Newton's Laws apply to physicists when doing calculations on a small scale. Thus, this use of race has been defined as race as Newtonian Physics (37).

Another approach to the concept of race is the common thought among forensic anthropologists that race is a necessary evil. The idea that there is no biological category for race, but the methods employed by forensic anthropologists give an 80-90% accuracy rate when the

methods are employed. Norman Sauer exemplifies this belief. Sauer argues that while there is more diversity within a group than between a group, there is the ability to determine an individual's classification. This classification can be done by employing many theories that often help give a biological profile of any unidentified remains. Thus, Sauer and other forensic anthropologists see this concept of race as a necessary evil, an essential aspect of returning unidentified remains to their loved ones.

While the necessary evil concept holds significant ground in forensic anthropology, others counter this argument of race being non-existent and race holds no utility in the realms of forensic anthropology. Smay and Armelagos argue that there are potentially numerous misidentified remains; this can seriously hamper or destroy any chance of identifying the remains. In some instances, misidentification has been discovered years later, allowing for proper identification (39). This argument has grown traction recently, initially placed in the spotlight at the 2020 American Association for Forensic Science (AAFS) Conference. In his oral presentation, Justin Maiers highlighted the dangers of ancestral estimation and the limitations of attempting to place individuals into categories based on a social construct (40). Racial tension converging to fever-pitched in the summer of 2020 after police brutality against the African American community rose. The subsequent uprising perpetuated a response from many within the forensic anthropological community, with DiGangi and Bethard promoting the concept of ignoring ancestral estimation (41). The premise of the argument was that policing authorities used ancestral estimation to ignore certain cases, specifically minority cases, and instead promote the needs of European American cases. Arguing, many of the unidentified remains are unidentified in part by labeling them with an ancestral category. Bethard and DiGangi further discuss how no determined relationship between hereditability and nonmetric skeletal traits has

occurred. Several researchers have countered this argument (42), calling for an evaluation of the methods. However, the call to remove ancestry has gained traction in the last few months for many forensic anthropologists.

In recent years, there has been a shift in terminology. The AAA and biological anthropologists further supported this push initiated with Montagu in the 1950s to tackle race problems. The issue was that there is still a significant need from the medicolegal field to rely on race or ancestry to define part of the biological profile. Those dealing with finding answers to what happened to these remains and who they are are not interested in a philosophical debate surrounding ancestry. Ross proposed addressing the definition of ancestry, suggesting a paradigm shift shedding the term's social construct (43). She identified that ancestry had limited bases on populations and how they have adapted to their environment and were more in tune with a European American ethnocentric view on ancestry—highlighting that populations should be defined based on clines, not loose interpretations. Ross argues that individuals should affiliate with populations with similar documented traits defining this as population affiliation.

Nevertheless, as academics, it is imperative to find effective methods to describe an individual's biological profile. Not only must we consider the scientific process but the social and cultural ramifications that these methods will incur. Thus, we have to rely on the tools at our disposal; understanding those tools and defining their strengths and weaknesses is critical in minimizing systems and user error.

CHAPTER THREE:

METHODS AND MATERIALS

We conduct multiple tests within 3D-ID and Fordisc 3.1 utilizing craniometrics data for n=36 positively identified individuals. We also exploit n=26 possible reference populations to interpret population affiliations estimated by the computer programs. While this study is not a direct comparison of Fordisc 3.1 and 3D-ID, we analyze how these programs can evaluate population affiliation. The desire is to outline which strategies yield higher accuracy for estimating population affiliation for unknown cases.

The cohort is a contemporary sample consisting of individuals whose birth and death occurred in the United States (2010-2020), including n=23 males and n=13 females. The age ranged of the cohort within the study at the time of death was 10-97, three individuals under 18 (ages 10,13 and 17), and four elderly individuals over 70. We used solved medicolegal cases from the sample; this included the juvenile and elderly cases mentioned above for estimating population affiliation through Fordisc 3.1 and 3D-ID for identification purposes.

The sample represents solved cases of previously unknown, unidentified remains analyzed by Dr. Erin Kimmerle and graduate student researchers working within the Forensic Anthropology Laboratory at the Florida Institute of Forensic Anthropology and Applied Science (USF-FAL) at the University of South Florida. The cases were analyzed using the Forensic Anthropology Laboratory (USF-FAL) protocols for ancestry estimation (44) that rely on metric analyses utilizing Fordisc 3.1 and 3D-ID. Following the analyses, we created facial

reconstructions for public engagement to aid in identification. The average of phenotypical variations for various populations, in part, assisted in creating facial approximations for many of the cases. Through DNA testing, identification for all individuals occurred. Each case had antemortem data regarding ancestry and family background obtained from the Medical Examiners records, law enforcement records, or court documents post-identification. The original analyses were "blind" because the remains were skeletonized at the time of analysis, with no prior identification of the victims available. Therefore, there is an exploration of the methods in this applied research, determining its accuracy. It also serves as a model for lab quality control.

Samples for this study were selected to fit the following two criteria: 1) The victim was initially unidentified but was later positively identified to evaluate known population history. 2) Linear and 3D coordinate craniometric data were available to test Fordisc 3.1 and 3D-ID methods for population affiliation.

Since Fordisc 3.1 and 3D-ID consist of different reference populations, the populations were organized into geographical regions standardized for comparison and mainly following the model established in Fordisc 3.1. For example, African Americans, referred to as "American Black" in Fordisc 3.1, or other African groups in 3D-ID were considered *African* ancestry (Table 2). European Americans, referred to as "White American" in Fordisc 3.1 and differentiated in 3D-ID either as "European" groups or "European American" were considered *European* ancestry (Table 2). Hispanic was considered any group from the Americas with Spanish first contact. However, Native Americans would classify as a population if any cases were present; none were. While there were no consistent Asian populations represented between the two programs, one individual in the sample had Asian Ancestry and fell into the Admixture (Admix) category.

The term Admix, or admixture, and Hispanic are widely used and highly problematic. "Admixture" is a common term used within forensic anthropology, denoting individuals with a combined ancestral heritage of human populations long separated prior to exploration and colonization (45). "Hispanic" is a catch-all term referring to populations spread throughout North America, the Caribbean, Central, and South America. The term does not represent one cultural nor biological population; instead, this concept has persisted as simply an ancestral nomenclature. This expansive group and lack of data surrounding the group have contributed to inaccuracies in this population's placement affinity (46–48).

In total, five individuals within the study fell into the Admix category, based on their parent's reported ancestry: one Asian/European, one African/Hispanic, one European/Hispanic, and two African/European (Table 2). Within the study, sex was known and did not act as a discriminatory factor, assigning appropriate biological sex to each case through known data. Four tests were each performed in Fordisc 3.1 and 3D-ID for the complete cohort. The program uses the mean scores of reference measurements from all selected ancestral groups to create appropriate centroids based on those measurements. The more reference samples used within the model, the less precise the results will be (20). Several measurements were greater than two standard deviations. These measurements' accuracy could not be determined, resulting in their removal due to potential errors (refer to descriptions below). The computer uses Posterior probabilities (*pp*) to determine how close an individual compares to each of the centroids for all selected reference populations. The practitioner selects which populations to incorporate in the analysis. The typicality (*typ*) assesses how likely the case is to belong to the cohort that compliments the centroid (25). The *typ* is based on D^2 and is further used to estimate affiliation with any group/cohort in the ancestral estimation (47,49).

Table 2. Conversion of Fordisc 3.1 & 3D-ID groups for comparison

Antemortem Data	Fordisc Population	3D-ID Population	Synthesized Label
Black	American Blacks	African African American East African West African	African
White	American White	European American European Central European Eastern European Southeastern European Southwestern	European
Hispanic	Hispanic	Circumcaribbean Colombian Mesoamerican South American	Hispanic
White/Hispanic	American White Hispanic	European American European Central European Eastern European Southeastern European Southwestern Circumcaribbean Colombian Mesoamerican South American	Admix
Asian/White	American White Chinese Japanese Vietnamese	European American European Central European Eastern European Southeastern European Southwestern East Asian	Admix
Black/Hispanic	Black American Hispanic	African African American East African West African European American European Central European Eastern European Southeastern European Southwestern	Admix
White/Black	Black American White American	African African American East African West African European American European Central European Eastern European Southeastern European Southwestern	Admix

Fordisc 3.1

The four tests in Fordisc 3.1 were as follows:

1. **Test 1: A Shotgun Approach.** Test 1 used the total sample (n=36) and all accessible populations (n=8), using 26 measurements to demonstrate a broad understanding of affiliation with the populations found in the FDB.
2. **Test 2: Altering the Measurements Included.** Test 2 utilized the entire sample (n=36) and all accessible populations. However, the removal of Biasterionic Breadth (ASB) and Zygomaxillary Breadth (ZMB) occurred as these measurements are known to reduce the sample size within FDB (20). Other measurements removed ≥ two standard deviations from the mean in Test 1; this occurred in 10 cases. The specific measurements removed for each case are discussed in the results when significant.
3. **Test 3: Incorporating Posterior Probabilities**. For Test 3, the removal of populations occurred when the *pp* ≤0.001 from Test 1. Testing was unavailable for nine cases due to the removal of specific populations, either because 1) No population had a *pp* ≤0.001, or 2) only one population had a *pp*>0.001. The exclusion of Vietnamese males and American Indian males occurred nine times; the exclusion of Japanese females occurred eight times; the exclusion of Guatemalan males and American Indian females occurred seven times; the exclusion of Chinese males and Japanese males occurred five times. The following populations were all removed once: Hispanic males, American White males, American Black males, Hispanic females, American White females, and American Black females, thereby reducing the pool of possible matches, leading to a more accurate result.
4. **Test 4: Removing both Outlying Measurements and Atypical Reference Populations.** The removal of populations occurred in Test 4 of populations with *pp*

≤0.001 from Test 2. The exclusion of measurements also occurred when they were greater than two standard deviations from the mean. The removal of the same reference populations from Test 3 also occurred in this test.

3D-ID

Akin to the population selection criteria for Fordisc 3.1, four tests were performed in 3D-ID, utilizing Generalized Procrustes Analysis (GPA) to compare the skulls' shape and size (21,27). By analyzing multivariate measurements, the program can estimate the best fit centroid for the specific case giving an estimation of both sex and ancestry. The computer program utilizes posterior probabilities (*pp)* and typicalities (*typ)* to estimate the affiliation a specific case has with an ancestral group, listing potential population cohesion (49). Following the recommendations of Ross and Slice (27), the tests followed specific criteria for group selection. The only test that did not meet the stipulations outlined by Ross and Slice was Test 4, which combined the variable criteria found in Tests 2 & 3 for this study. Due to a constant measurement error, the removal of one case occurred from the cohort. Rectification of the errors was unobtainable due to the inability to re-evaluate measurements. Thus, the removal of an established European ancestry case occurred. The sample for all 3D-ID tests was n=35. At the time of testing, the identified ancestry was unknown, and each test was run blind. Again, all tests employed appropriate sex estimation with the following parameters:

1. **Test 1: The Shotgun Approach**. Test 1 included all recommended measurements (n=23) for a generalized test, as outlined by Slice and Ross (27), using all of the available reference groups (n=19) as defined in Table 3.

2. **Test 2: Measurement Selection Based on European Populations.** Slice and Ross (27) suggest specific reference samples for Brazilian and European cases and the exclusion of the following measurements when using this test: Ectoconchion left and right (ectl, ectr), Lower Orbital Border left, and Upper Orbital Border left (obhi, obhs), Opisthion (ops). The inclusion following measurements occurred from Test 1: ecml, ecmr, mastl, mastr (Table 3).
3. **Test 3: Measurement Selection Based on MesoAmerican Populations.** For the estimation of MesoAmerican populations, Slice and Ross (27) include the following reference samples: Mayan, Indigenous, and European populations. Therefore, the exclusion of the following measurements occurred for this test: Frontomalare Temporale left and right (fmtl, fmtr), Subspinale (ssp), Zygomaxilare left and right (zygoml, zygomr), Zygoorbitale left and right (zygool, zygr), ecml, ecmr, obhi, obhs, obhsr, obhir, mastl, and mastr (Table 3).
4. **Test 4: Combination of Measurement Selection Based on MesoAmerican, European, and Brazilian Reference Samples.** Test 4 used the following measurements: Asterion left and right (astl, astr), Basion (bas), Bregma (brg), Dacryon left (dacl), Glabella (glb), Nasion (nas), Zygion left (zygl), fmal, fmar, proHEST, and zygr (Table 3).

Table 3. Identifies all Landmarks utilized for Tests 1-4.

Landmark Name	Landmark Abbreviation	Test 1 General	Test 2 European and Brazilian	Test 3 Mesoamerican	Test 4: Derivative of Test 2 & 3
Asterion Left	astl	✓	✓	✓	✓
Asterion Right	astr	✓	✓	✓	✓
Basion	bas	✓	✓	✓	✓
Bregma	brg	✓	✓	✓	✓
Dacryon Left	dacl	✓	✓	✓	✓
Ectomolare Left	ecml		✓		
Ectomolare Right	ecmr		✓		
Ectoconchion Left	ectl	✓		✓	
Ectoconchion Right	ectr	✓		✓	
Frontomalare Anterior Left	fmal	✓	✓	✓	✓
Frontomalare Anterior Right	fmar	✓	✓	✓	✓
Frontomalare Temporale Left	fmtl	✓	✓		
Frontomalare Temporale Right	fmtr	✓	✓		
Glabella	glb	✓	✓	✓	✓
Lambda	lam	✓	✓	✓	✓
Mastoideale Left	mastl		✓		
Mastoideale Right	mastr		✓		
Nasion	nas	✓	✓	✓	✓
Lower Orbital Border Left	obhi	✓		✓	
Upper Oribital Border Left	obhs	✓		✓	
Opisthion	ops	✓		✓	
Prosthion-Howells Estimated	proHEST		✓	✓	✓
Subspinale	ssp	✓			
Nasomaxillary Suture Pinch Left	wnbl			✓	
Nasomaxillary Suture Pinch Right	wnbr			✓	
Zygion Left	zygl	✓	✓	✓	✓
Zygomaxilare Left	zygoml	✓	✓		
Zygomaxilare Right	zygomr	✓	✓		
Zygoorbitale Left	zygool	✓	✓		
Zygion Right	zygr	✓	✓	✓	✓
Total number of measurements used		23	22	20	13

Comparison

Each test's overall performance is analyzed to evaluate how the populations and measurement variation affected the tests' outcome. Comparison occurred of Fordisc 3.1 tests to establish potential relationships within the four tests, using Fisher's exact test. For 3D-ID, the application of identical statistical analysis occurred. The final step was applied when we compared correct estimations between Fordisc 3.1 and 3D-ID utilizing the same methods. We could not perform a parallel analysis on all tests due to the different parameters found within the tests. Therefore, a comparison of all Fordisc 3.1 Tests against 3D-ID Tests 1 & 4 utilizing

Fisher's exact test occurred. When the result was $p \leq 0.05$, it was considered statistically significant and demonstrated differences between the compared tests.

We then analyzed significant differences between group classifications for each test. However, due to the sample size difference, the *z-score* test for two populations was utilized to analyze small and varying sample sizes. We compared like groups with each other; all the European groups were compared for the Fordisc 3.1 tests, as were the African groups and Admixture groups. There were no comparisons for Hispanic groups, as the sample size was too small. For 3D-ID tests, the same comparisons occurred. The final comparison was assessing Fordisc 3.1 and 3D-ID. However, Test 2 for Fordisc 3.1 was not directly comparable with 3D-ID Test 2. Thus, we compared all the Fordisc 3.1 tests with two 3D-ID tests, the two tests not designed to focus on specific ancestral groups. Therefore, evaluating 3D-ID Test 1 & 4 against all Fordisc 3.1 tests ensued. Some individuals' exclusion in Fordisc 3.1 Tests 2 & 3 happened as they did not meet inclusion parameters. The test sizes varied and further reduced the number of cases. These inconsistent and small sizes led to concerns for consistency parameters resulting in the use of the *z-score* test for two populations, which utilized the following formula:

$$z = \frac{[(\hat{p}1 - \hat{p}2) - 0]}{\sqrt{[\hat{p}(1-\hat{p})\left(\frac{1}{n1} + \frac{1}{n2}\right)]}}$$

In this equation, $\hat{p}$ represents the number of correct estimations within the ancestral cohort for the specific test. The *n* represents the sample size of the ancestral cohort for that test as there was a different number of accurate estimations between the tests when comparing ancestral cohorts. We then assessed any significant difference in the test's performance. After establishing the *z-score*, we determined the *p*-value. When comparing two like ancestral cohorts, a significant

difference occurred with results of $p \leq 0.05$. We compared each ancestral cohort against each of the Fordisc 3.1 tests. We repeated this process for 3D-ID. Both computer programs' comparison occurred as all ancestral cohorts within all Fordisc 3.1 tests against all ancestral cohorts within 3D-ID Tests 1 & 4. Due to the sample size, no analysis occurred of the Hispanic cohort. Thus the exclusion of this cohort occurred for these tests.

The *pp* and *typ* for the correct results were analyzed separately. To analyze the results accurately, we divided them by computer program and test. The generation of descriptive statistics occurred utilizing SPSS v.26 for the *pp* and *typ* in Fordisc 3.1 and 3D-ID, which indicated potential trends in the data distribution. We plotted *pp* and *typ* together to establish other trends once they were analyzed individually. We analyzed this to outline how confident in the results a practitioner utilizing the data could be, as the computer program estimating ancestry is only helpful when there is high confidence in the findings. Thus, higher *pp* and *typ* should encourage a practitioner to be comfortable with their potential estimations.

Since it is up to every investigator to evaluate either program's results and use their interpretation to make an ancestry estimation, we graphed the posterior probabilities and typicalities in Microsoft Excel for each Test. Each Test was further broken down into the correct number of classifications and plotted to utilize the *typ* over *pp* results. The population affiliations were assigned a color: red for European, blue for African, orange for Hispanic, and green for admixture. We established ancestral grouping and the estimated populations created by Fordisc 3.1 or 3D-ID for the misclassified cases. The *pp/typ* is divided into four quadrants: Quadrant 1 – *pp/typ*<0.5; Quadrant 2 – *pp*>0.5, *typ*<0.5; Quadrant 3 – *pp*<0.5, *typ*>0.5; Quadrant 4 – *pp/typ*>0.5.

CHAPTER FOUR:

RESULTS

Fordisc 3.1

Test Accuracy and Correct Classifications

The overall success for correct classifications in Fordisc 3.1 for all four tests was over 80.0% (Figure 1). As expected, there was a high degree of variation within each cohort for each test, and the frequency of correct classifications depended on the population reference samples included. Test 1 results show correct classification was highest for the African cohort (100%), followed by Europeans (83.3%), Admixed cohort (80.0%), and then Hispanics (0.0%). The overall percentage for Test 1 was 80.6%.

Test 2 had an overall accuracy of 83.3% and was higher than that of Test 1 for Europeans, but otherwise followed the same pattern as Test 1: African 100%, European 87.5%, Admixture 80.0%, and finally Hispanic 0.0%.

The overall accuracy demonstrated in Test 3 was slightly less accurate at 81.4%, with the classifications following the same pattern: African 100%, European 87.5%, Admixture 80.0%, and once again Hispanic 0.0%. The final test, Test 4, produced an overall accuracy rate of 82.7%. The classification groups reached the following accuracy levels, with the African cohort achieving 100%, European 88.9%, Admixture 60.0%, and Hispanic 0.0%. Note that the Admix

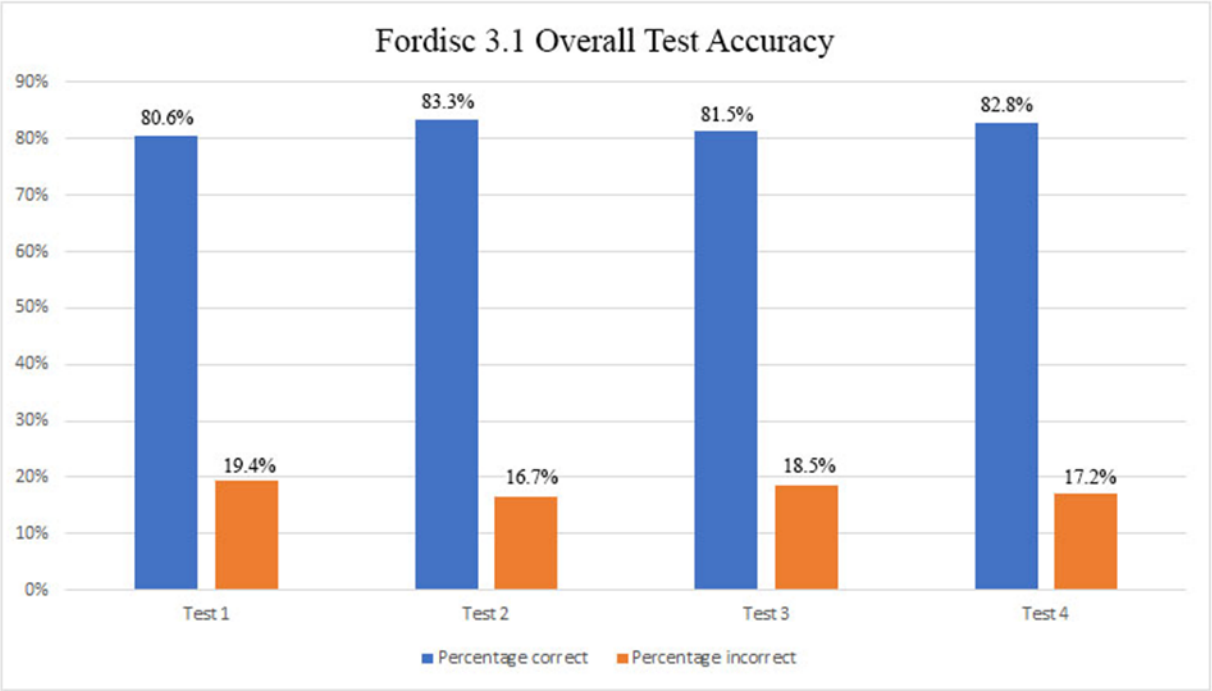

Fig1. Classification Accuracy Estimated through Fordisc 3.1

group performed lowest in test four, no Hispanic cohorts correctly classified for any of the Fordisc 3.1 tests.

Fishers Exact Test & Z-Score Test for Two Populations

Fisher's exact test showed no significant difference ($p \leq 0.05$) in the performance between any of the four Fordisc 3.1 tests. When the classification groups were compared for each test, using the z-scores for two populations, there was no significant difference ($p \leq 0.05$) for the European cohort, African cohort, or Admixture cohort. All the *p*-values for each of the compared groups were $p > 0.49$.

Posterior Probabilities

Using Fordisc 3.1, the four tests show similar results for the *pp* among correct classifications (Table 4), the range for the tests extends from 0.6790 to 0.7540. Fordisc 3.1 Test 1 demonstrated a left skew of -1.8110 and a range of 0.7080. The kurtosis is -0.6090, and this results in a platykurtic and left-skewed distribution for Test 1. There is a left skew in all the other

tests for Fordisc 3.1, along with negative kurtosis, illustrating that each of the Fordisc 3.1 test distributions is platykurtic and left-skewed, having more *p*-values closer to 0 than 1 (Table 4).

Typicalities

The *typ* observed in Fordisc 3.1 tests shows a broader range than *pp* in each of the four tests (Table 5). There is negative kurtosis in all four tests for *typ* for Fordisc 3.1, giving a consistent platykurtic distribution. All skews are narrower than *pp* giving a more normal distribution for all tests. Tests 1, 3, and 4 have a slight right skew, not seen in Fordisc 3.1 *pp* results (Table 5).

Understanding Classification

Test 1 for Fordisc 3.1 had 80.6% correctly estimated cases (29/36). Table 5 shows the number of cases relative to their posterior probability and typicality; in this test Quadrant 2 had the highest percentage (48.2%), with Quadrant 3 having the lowest percentage (3.4%) (Table 6). Misclassified estimations demonstrated far more interesting results (n=7) (Figure 2). Figure 2 shows individuals misclassified, including one individual from a European ancestry classified highest with the Chinese reference group. The two Hispanic cases were classified as European, shown in Quadrant 4. Two individuals could not be classified, whose ancestry was European, and both fell within Quadrant 2.

Table 4. Posterior Probability Descriptive Statistics for Fordisc 3.1 Tests 1-4.

		Correct Posterior Probabilities								
Test	**n**	**Mean Score**	**Mean Error**	**Std Deviation**	**Range**	**Interquartile range**	**Skewness Score**	**Skewness error**	**Kurtosis Score**	**Kurtosis error**
Test 1 Fordisc	30	0.7755	0.0401	0.2198	0.7080	0.3290	-0.8110	0.4270	-0.6090	0.8330
Test 2 Fordisc	30	0.7030	0.0417	0.2283	0.7540	0.3700	-0.3430	0.4270	-0.9180	0.8330
Test 3 Fordisc	22	0.7794	0.0470	0.2204	0.6830	0.3410	-0.7340	0.4910	-0.7650	0.9530
Test 4 Fordisc	24	0.7752	0.0412	0.2017	0.6790	0.3720	-0.4050	0.4720	-0.9290	0.9180

Table 5. Typicality Descriptive Statistics for Fordisc 3.1 Tests 1-4.

		Correct Typicality								
Test	**n**	**Mean Score**	**Mean Error**	**Std Deviation**	**Range**	**Interquartile range**	**Skewness Score**	**Skewness error**	**Kurtosis Score**	**Kurtosis error**
Test 1 Fordisc	30	0.3868	0.0519	0.2841	0.9580	0.5320	0.3170	0.4270	-0.8970	0.8330
Test 2 Fordisc	30	0.5047	0.0479	0.2621	0.9430	0.4410	-0.0600	0.4270	-0.7170	0.8330
Test 3 Fordisc	22	0.3529	0.0636	0.2981	0.9350	0.4790	0.4410	0.4910	-0.8860	0.9530
Test 4 Fordisc	24	0.5138	0.0559	0.2734	0.9300	0.4440	0.0580	0.4720	-0.7670	0.9180

Test 2 for Fordisc 3.1 showed 83.3% correct classifications (30/36). Quadrant 4 gave the highest percentage with 50.0% of all correct classifications (Table 6), while both Quadrant 1 and Quadrant 4 had the lowest percentage (6.7%). Figure 3 highlights the six misclassified cases. Misclassification occurred for both Hispanics, identified as European. Misclassification occurred for three European cases, all estimated with origins from the Americas (Figure 3). Test 2 performed better than any other tests and had a high percentage of cases within Quadrant 4.

Test 3 for Fordisc 3.1 showed 81.5% correct classifications (22/27) (Table 6). The highest percentage for correct estimations was Quadrant 2 (54.5%), and the lowest was Quadrant 3 (0.0%). Figure 4 illustrates five misclassifications. The misclassification of the entire Hispanic cohort occurred again, with both estimated as European with strong *pp*. Test 3 performed the weakest as most of the correctly estimated cases fell in Quadrants 1 and Quadrants 2, resulting in low *typ* and the misclassifications mostly fell in high *pp* quadrants.

Test 4 for Fordisc 3.1 showed 82.7% correct classifications (24/29). Quadrant 4 had the largest percentage (50.1%), with Quadrant 3 giving the lowest percentage (0.0%) (Table 6). There were five misclassified cases, four of which were in a quadrant with high *pp* (Figure 5). Misclassification occurred for two admixture cases; both labeled as Hispanic and Native American. Only one Hispanic case was analyzed in this test, again classified as likely European.

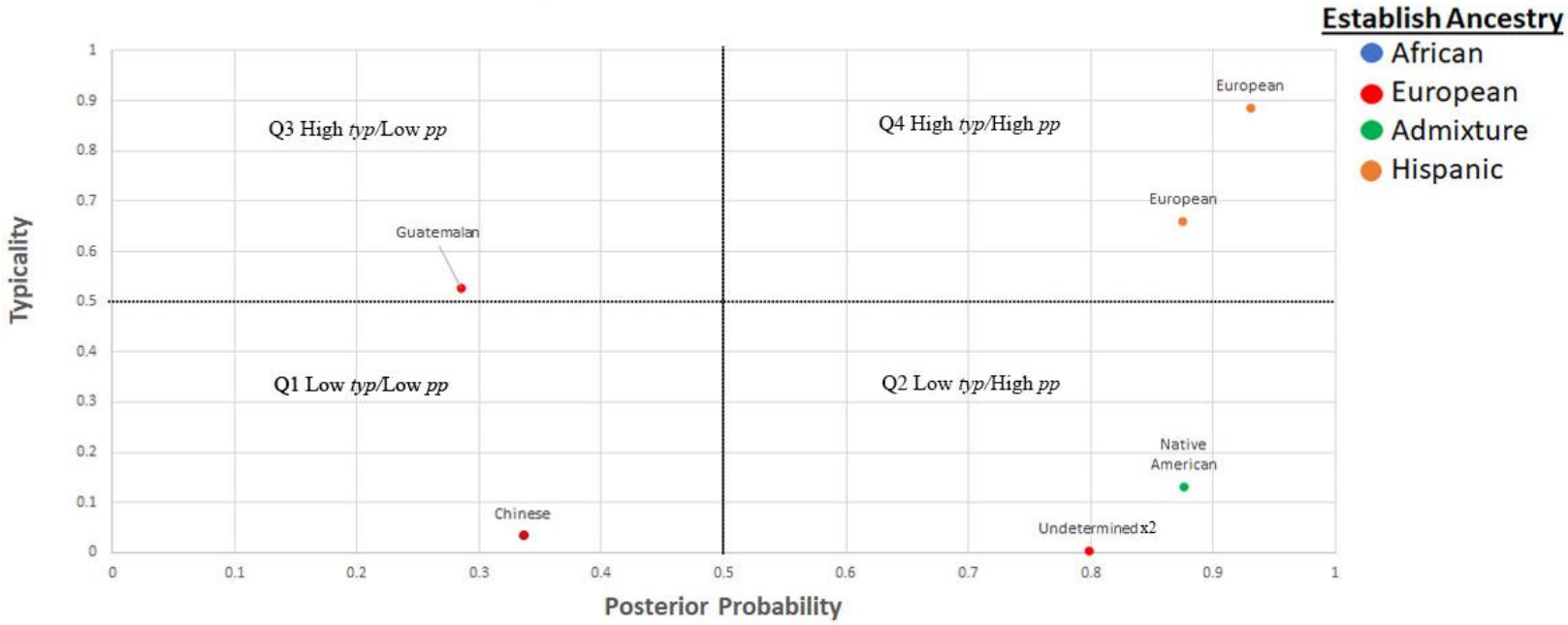

Fig 2. Demonstrates the distribution of incorrect estimated ancestral cases for *pp* and *typ* with Fordisc 3.1 Test 1.
The establish ancestry, defined by DNA results is listed as establish ancestry. The written notations identify estimated ancestry by Fordisc Test 1

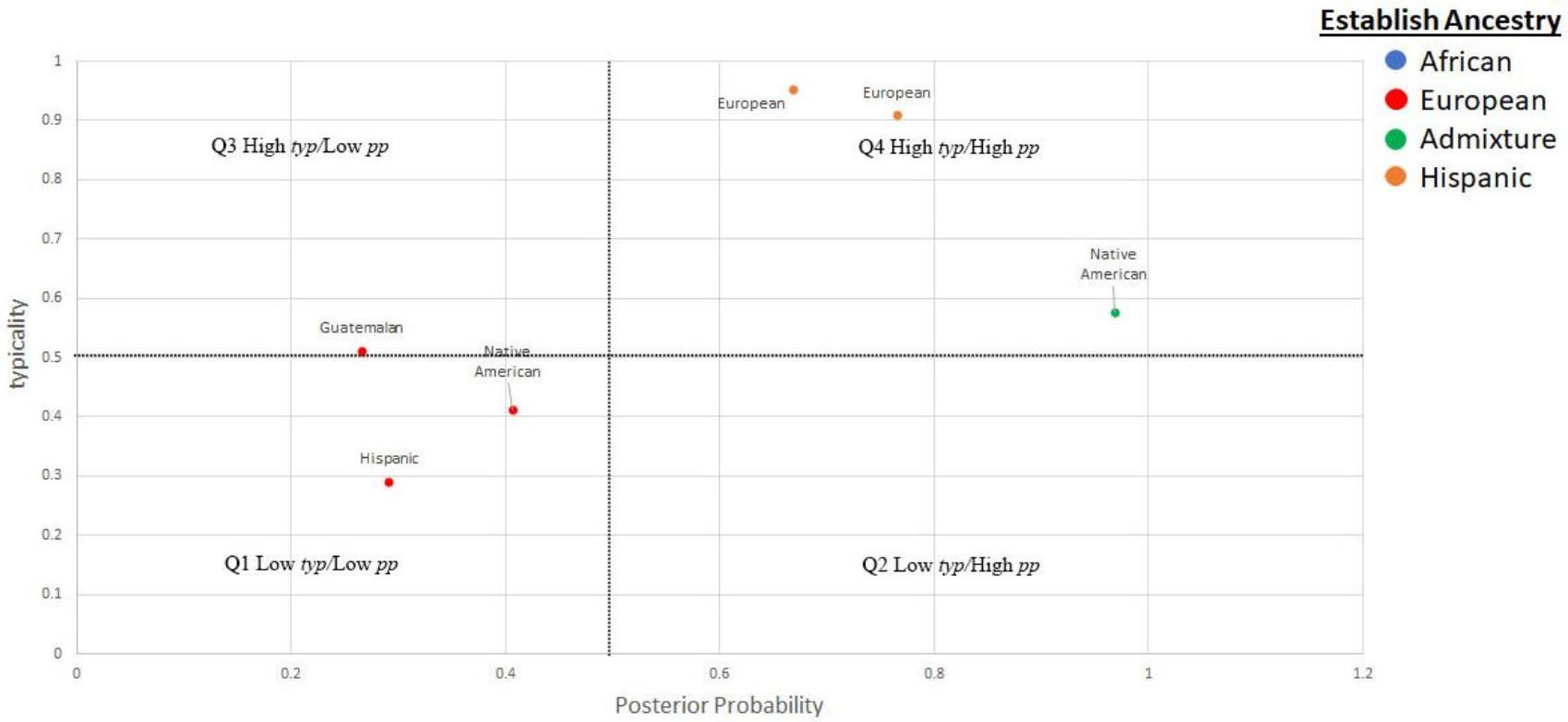

Fig 3. Demonstrates the distribution of incorrect estimated ancestral cases for *pp* and *typ* with Fordisc 3.1 Test 2.
The establish ancestry, defined by DNA results is listed as establish ancestry. The written notations identify estimated ancestry by Fordisc Test 2.

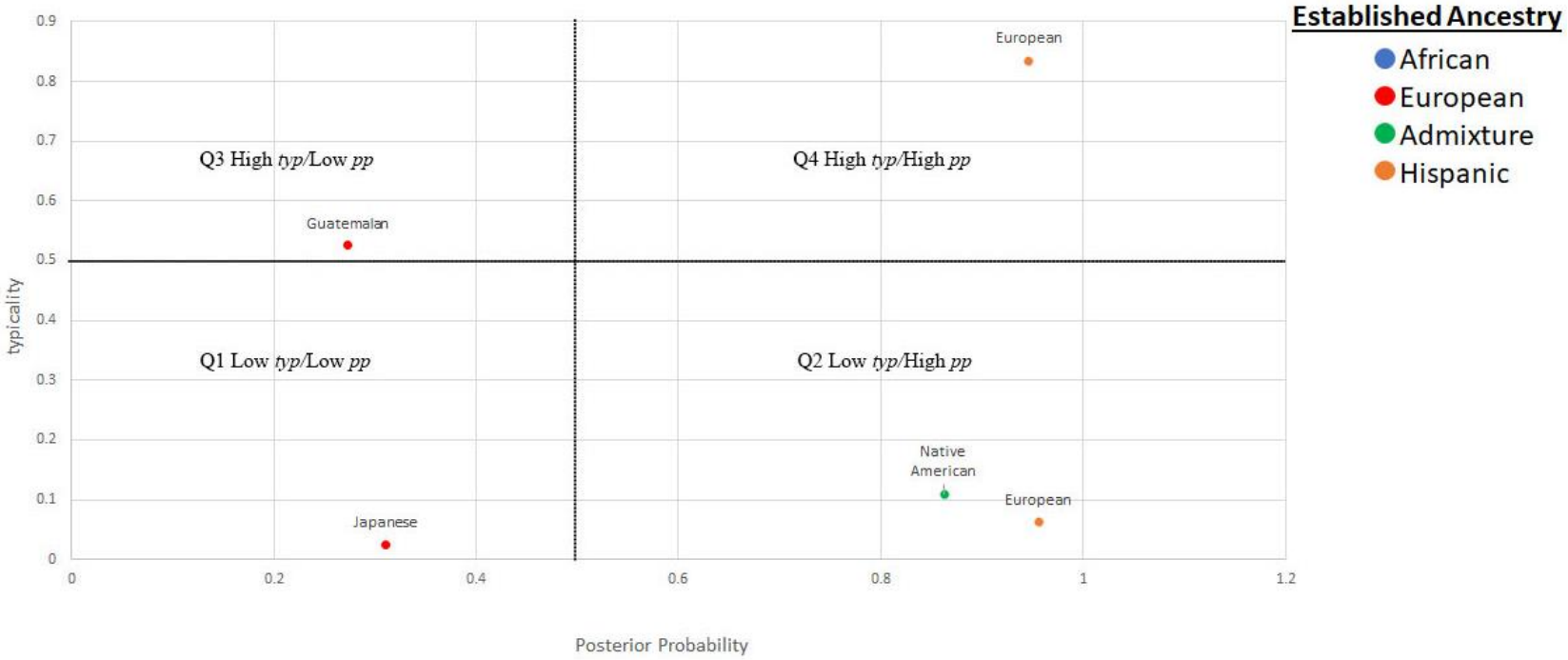

Fig 4. Demonstrates the distribution of incorrect estimated ancestral cases for *pp* and *typ* with Fordisc 3.1 Test 3.
The establish ancestry, defined by DNA results is listed as establish ancestry. The written notations identify estimated ancestry by Fordisc Test 3

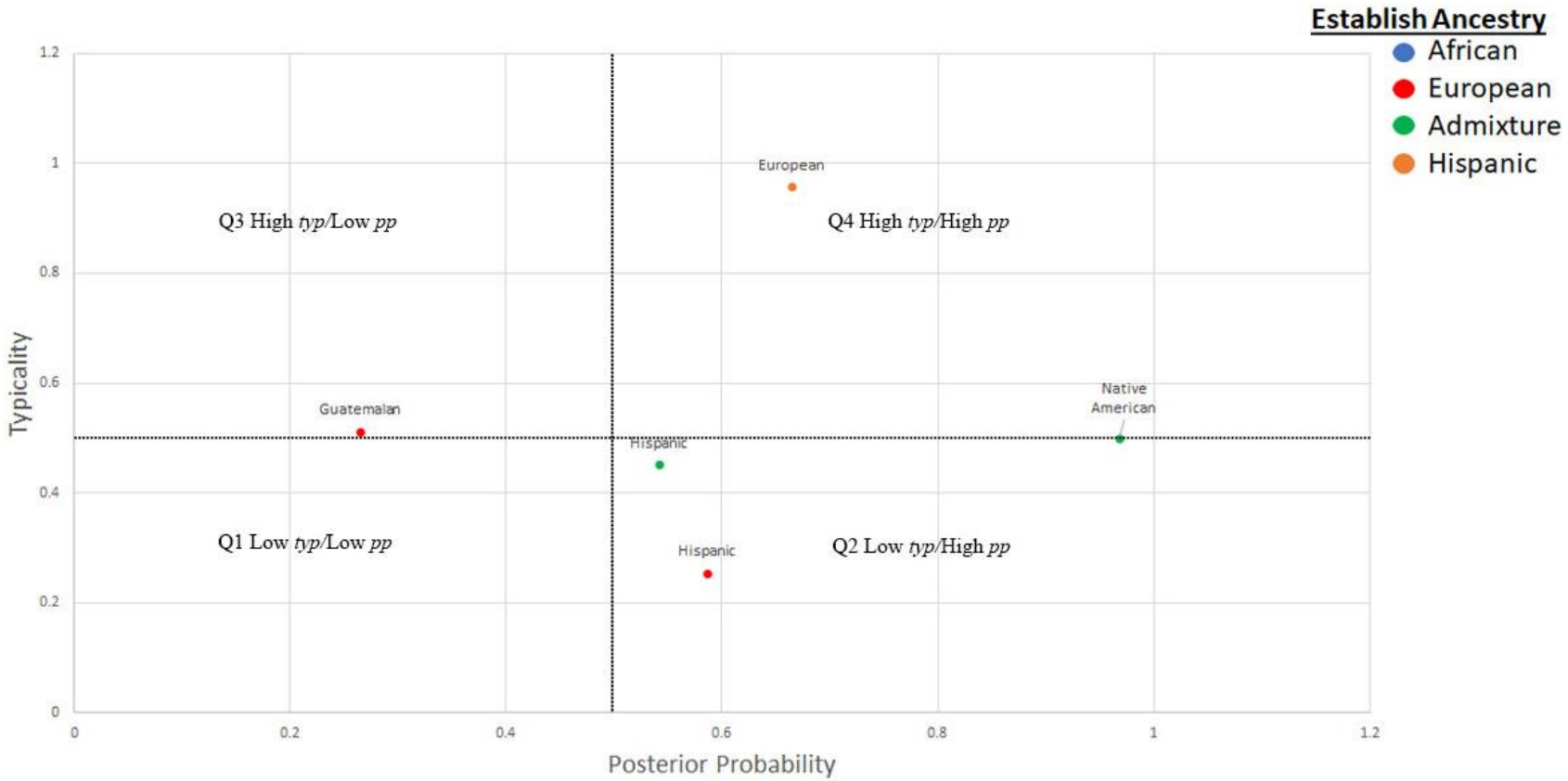

Fig 5. Demonstrates the distribution of incorrect estimated ancestral cases for *pp* and *typ* with Fordisc 3.1 Test 4.
The establish ancestry, defined by DNA results is listed as establish ancestry. The written notations identify estimated ancestry by Fordisc Test 4

Table 6. Fordisc 3.1 Correct Classifications by Quadrant.

		Fordisc Correct classifications (%)				Fordisc Incorrect classifications (%)			
		Test 1	**Test 2**	**Test 3**	**Test 4**	**Test 1**	**Test 2**	**Test 3**	**Test 4**
	European	10.3	6.7	9.1	4.2	14.3	33.3	20.0	0.0
	African	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Q1*	Hispanic	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Admixture	3.4	0.0	9.1	0.0	0.0	0.0	0.0	0.0
	Total	**13.7**	**6.7**	**18.2**	**4.2**	**14.3**	**33.3**	**20.0**	**0.0**
	European	34.5	26.7	36.4	33.3	28.6	0.0	0.0	20.0
	African	10.3	0.0	13.6	4.2	0.0	0.0	0.0	0.0
Q2*	Hispanic	0.0	0.0	0.0	0.0	0.0	0.0	20.0	0.0
	Admixture	3.4	3.4	4.5	8.3	14.3	0.0	20.0	40.0
	Total	**48.2**	**30.1**	**54.5**	**45.8**	**42.9**	**0.0**	**40.0**	**60.0**
	European	3.4	6.7	0.0	0.0	14.3	16.7	20.0	20.0
	African	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Q3*	Hispanic	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Admixture	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Total	**3.4**	**6.7**	**0.0**	**0.0**	**14.3**	**16.7**	**20.0**	**20.0**
	European	20.7	26.7	18.1	29.2	0.0	0.0	0.0	0.0
	African	6.9	13.3	4.5	16.7	0.0	0.0	0.0	0.0
Q4*	Hispanic	0.0	0.0	0.0	0.0	28.6	33.3	20.0	20.0
	Admixture	6.9	10.0	4.5	4.2	0.0	16.7	0.0	0.0
	Total	**34.5**	**50.0**	**27.1**	**50.1**	**28.6**	**50.0**	**20.0**	**20.0**

**Q1 – pp/typ,0.5; Q2 – pp>0.5, typ<0.5; Q3 – pp<0.5, typ>0.5; Q4 – pp/typ>0.5*

3D-ID

Test Accuracy and Correct Classifications

3D-ID has an overall accuracy ranging between 48.6% to 74.3% (Figure 6). Test 1 total accuracy was 68.6% (24/35). However, there was a high variance between the different ancestral cohorts, with each group performing correctly to the following levels: Admix 80%, European 73.9%, African 60.0%, and Hispanic 0.0%.

Test 2 produced an overall accuracy rate of 65.7% (23/35). Each ancestral cohort varied for correct classifications at the following rates: Admixture 80.0%, European 69.6%, African 60.0%, and Hispanic 0.0%.

For Test 3, the overall accuracy was 48.6% (17/35). The frequency for correct classification were as follows: African 80.0%, Admixture 60.0%, European 43.4%, and Hispanic 0.0%.

For Test 4, the overall accuracy was 74.3% (26/35), the ancestral cohorts achieved the following correct classification rates: European 87.0%, African 80.0%, Hispanic 50.0%, and Admixture 20.0%. Test 4 was the only test to affiliate a Hispanic case to a Hispanic population accurately.

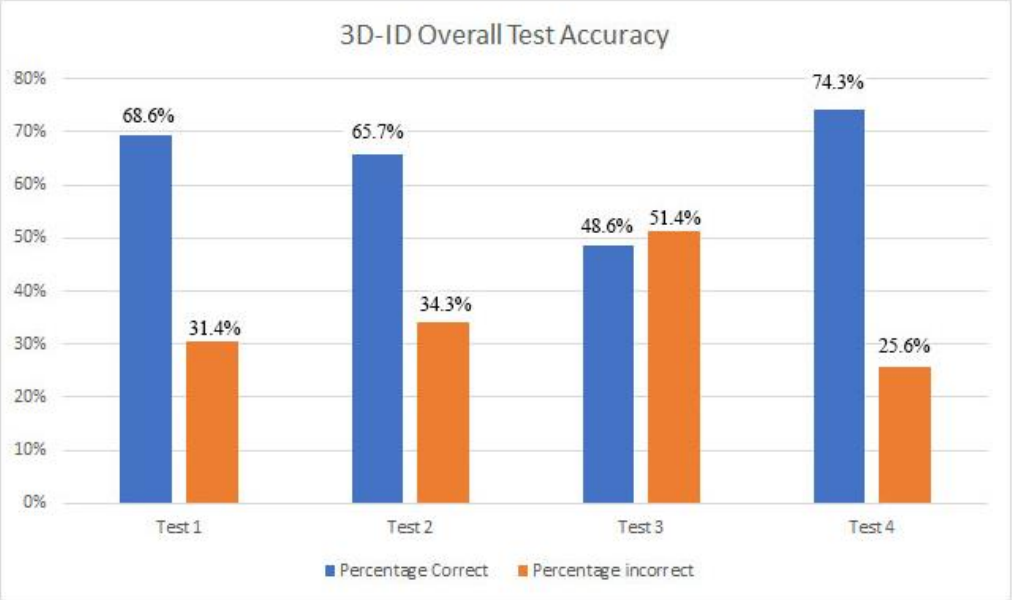

Fig 6. Classification Accuracy Estimated through 3D-ID

Fishers Exact Test & Z-Score Test for Two Populations

The Fisher's exact test comparing the 3D-ID tests demonstrated a statistically significant difference between Test 3 versus Test 4 (p=0.049). The *z-score* test for two populations also highlighted a statistically significant difference for Tests 3 versus Test 4 ($p=0.0271$). The *z-score* test for two Populations for European cohorts when comparing Test 1 versus Test 3 had a marginally significant difference ($p=0.057$), as did the Admixture groups between Test 1 versus Test 4, and Test 4 versus Test 2 ($p=0.057$ for both). No other comparisons within 3D-ID had any statistically significant results.

Posterior Probabilities

The *pp* for 3D-ID tests showed a range of 0.7999 (Test 4) to 0.9991 (Test 1) (Table 7). Tests 1-3 have negative skews, with both Tests 1 and 3 being over -1.0. As a result, many of the *pp* values are closer to a *p*-value of 0 than 1. However, Test 4 has a positive skew of 1.6480,

delivering *p*-values closer to 1 for *pp.* The kurtosis within the four tests shows a differing range of values. Test 2 and Test 3 have normal distributions, whereas Test 1 and Test 4 produce a leptokurtic distribution.

Typicalities

For all 3D-ID tests, the *typ* data shows a narrow range for correct classifications (Table 8). The lowest range, Test 4, is 0.9436, and the most extensive range, Test 1, is 0.9868 (Table 8). Test 3 (-0.4560 skew) and Test 4 (-0.1370 skew) have two tests with negative skews. Test 1 (0.9550 skew) and Test 2 (0.7630 skew) have positive skews. The estimations have negative kurtosis giving a platykurtic distribution. Test 1 has the closest to a normal distribution (-0.2000 kurtosis), and Test 3 had the most profound platykurtic distribution with a -1.4190 kurtosis.

Understanding Classification

We used A shotgun approach for Test 1, including all reference groups and possible measurements in the analysis. 68.8% of all cases were correctly classified (24/35) (Table 9). Quadrant 4 had the highest success rate at 54.2%, and Quadrant 3 had the lowest success rate at 0.0%. This test boasted a high *pp* throughout the test but varying *typ*. There were 11 misclassifications (Figure 7). The misclassification of both Hispanic cases occurred and estimated them European Americans. The misclassifications assignment appeared to be related to South American populations, with several European and African cohorts classified as Colombian.

Table 7. Posterior Probability Descriptive Statistics for 3D-ID Tests 1-4.

		Correct Posterior Probability								
Test	**n**	**Mean Score**	**Mean Error**	**Std Deviation**	**Range**	**Interquartile range**	**Skewness Score**	**Skewness error**	**Kurtosis Score**	**Kurtosis error**
Test 1 3D-ID	23	0.7327	0.0635	0.3046	0.9991	0.3803	-1.3080	0.4810	0.8280	0.9350
Test 2 3D-ID	21	0.6842	0.0540	0.2476	0.8149	0.3710	-0.5760	0.5010	-0.3900	0.9720
Test 3 3D-ID	17	0.7754	0.0609	0.2512	0.8175	0.3886	-1.1070	0.5500	0.2780	1.0630
Test 4 3D-ID	25	0.4347	0.0353	0.1766	0.7999	0.1837	1.6480	0.4640	3.4580	0.9020

Table 8. Typicality Descriptive Statistics for 3D-ID Tests 1-4.

		Correct Typicality								
Test	**n**	**Mean Score**	**Mean Error**	**Std Deviation**	**Range**	**Interquartile range**	**Skewness Score**	**Skewness error**	**Kurtosis Score**	**Kurtosis error**
Test 1 3D-ID	23	0.3135	0.0648	0.3109	0.9868	0.3790	0.9550	0.4810	-0.2000	0.9350
Test 2 3D-ID	20	0.3659	0.0779	0.3482	0.9800	0.5643	0.7630	0.5120	-0.8070	0.9920
Test 3 3D-ID	15	0.5426	0.9655	0.3739	0.9751	0.8593	-0.4560	0.5800	-1.4190	1.1210
Test 4 3D-ID	25	0.5554	0.0617	0.3085	0.9436	0.5371	-0.1370	0.4640	-1.3080	0.9020

Test 2 eliminated specific measurements to propagate the European reference samples found within 3D-ID. This test correctly classified 65.7% of all cases (23/35) (Table 9). There was an expectation that the European cohort would classify at a higher rate in this test. The European cohort achieved a 69.6% success rate, with two 3D-ID tests performing at a higher standard (Tests 1 and 4). Quadrant 2 had the highest correct classifications rate (52.1%), with Quadrant 3 producing the lowest rate (4.3%). There are 12 misclassified cases (Figure 8). Again, Hispanic cohorts are classified into European populations. All the misclassifications were either of European or Brazilian reference populations except for two European cases classified as Colombian. There was a lean in this test towards the European and Brazilian populations, possibly due to the landmarks utilized to exploit both Brazilian and European reference samples.

Test 3 utilized landmarks on the skull for analysis that exploited the MesoAmerican reference populations. For the total sample, the correct estimation was 48.6% (17/35) (Table 9). This test should favor individuals with MesoAmerican ancestry; however, no classifications of individuals with MesoAmerican or South American ancestry occurred; rather, they all aligned closest to the European groups. Quadrant 4 has the largest rate for correct classifications (52.9%), while Quadrant 3 shows no correct classifications. There were 18 misclassified cases (Figure 9). The misclassification of Hispanic cases estimated as European occurred in Quadrant 4. Four cases had pp=0.0 and typ=0.0. Such low results demonstrate a lack of confidence from 3D-ID in its analysis of these cases; this lack of confidence allows the practitioner to dismiss these results.

Test 4 selects specific landmarks to exploit European, MesoAmerican, and Brazilian reference populations (see methods section). The Hispanic cohort had the highest success rate than any other test within this study (50%). Overall, the success rate was 74.3% (26/35) (Table

8). While this was the highest correct classification for any of the 3D-ID Test, the correct estimations were found mainly in Quadrant 3 (42%), giving low *pp* but high *typ*. Figure 10 provides details about the nine misclassified cases, 66.6% of which fall into Quadrant 2. Only one case (Syrian, Quadrant 2) was not from a European population or populations found in the Americas. These results are consistent with the reference populations accentuated in this test.

Comparing Methods

We compared each of the four tests in Fordisc 3.1 and 3D-ID utilizing Fisher's exact test, and there were no significant differences in the model's performance ($p \leq 0.05$). These tests were analyzed by group classifications using *z-score*s for two populations for each set of samples. The European and African groups had statistically significant differences when compared to the tests performed within Fordisc 3.1. For Admix cohorts in 3D-ID Test 4 versus Admix cohorts in Fordisc 3.1 Tests 1-3, there were marginal differences observed; all these comparisons yielded a $p = 0.057$.

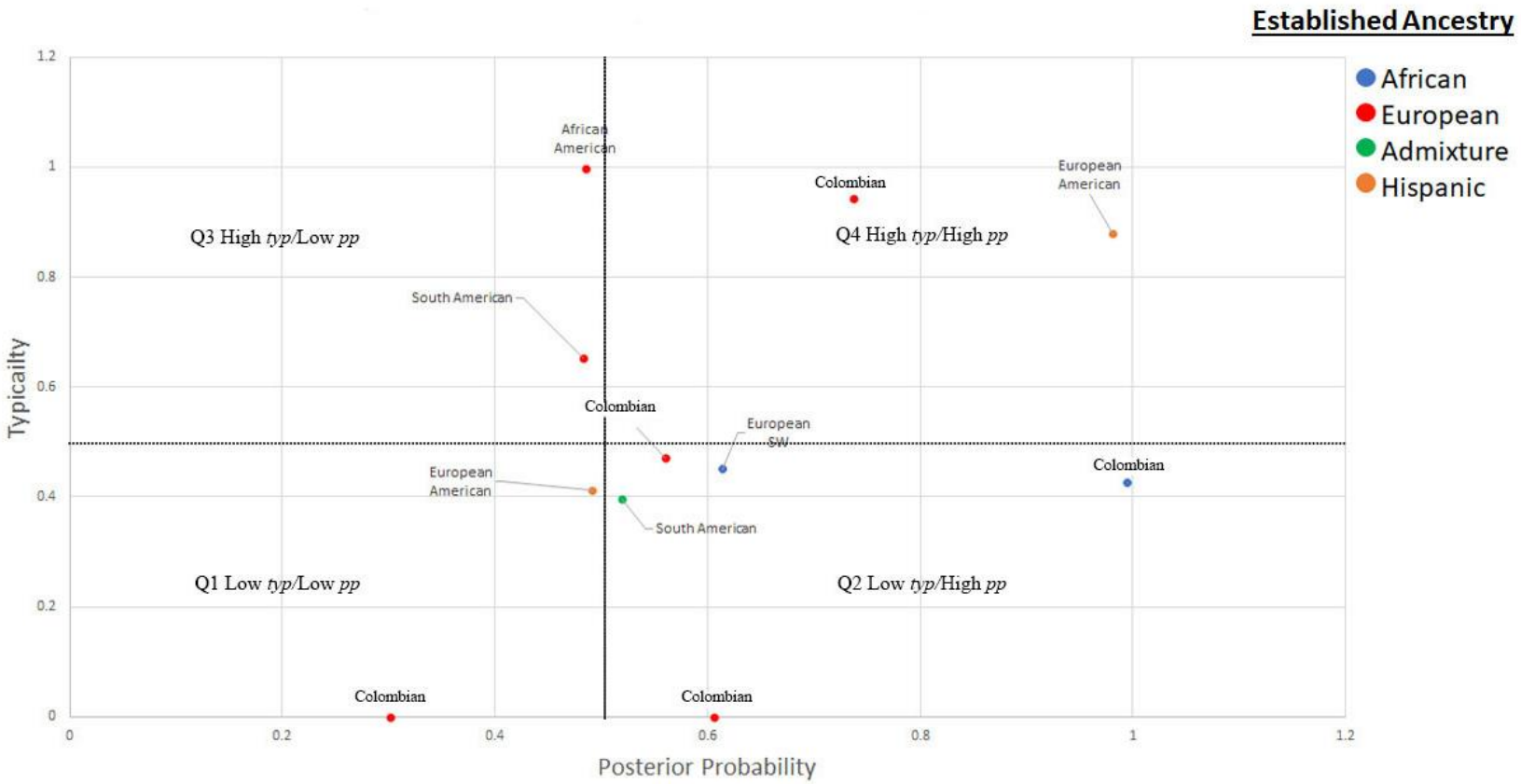

Fig 7. Demonstrates the distribution of incorrect estimated ancestral cases for *pp* and *typ* with 3D-ID Test 1.
The establish ancestry, defined by DNA results is listed as establish ancestry. The written notations identify estimated ancestry by 3D-ID Test 1

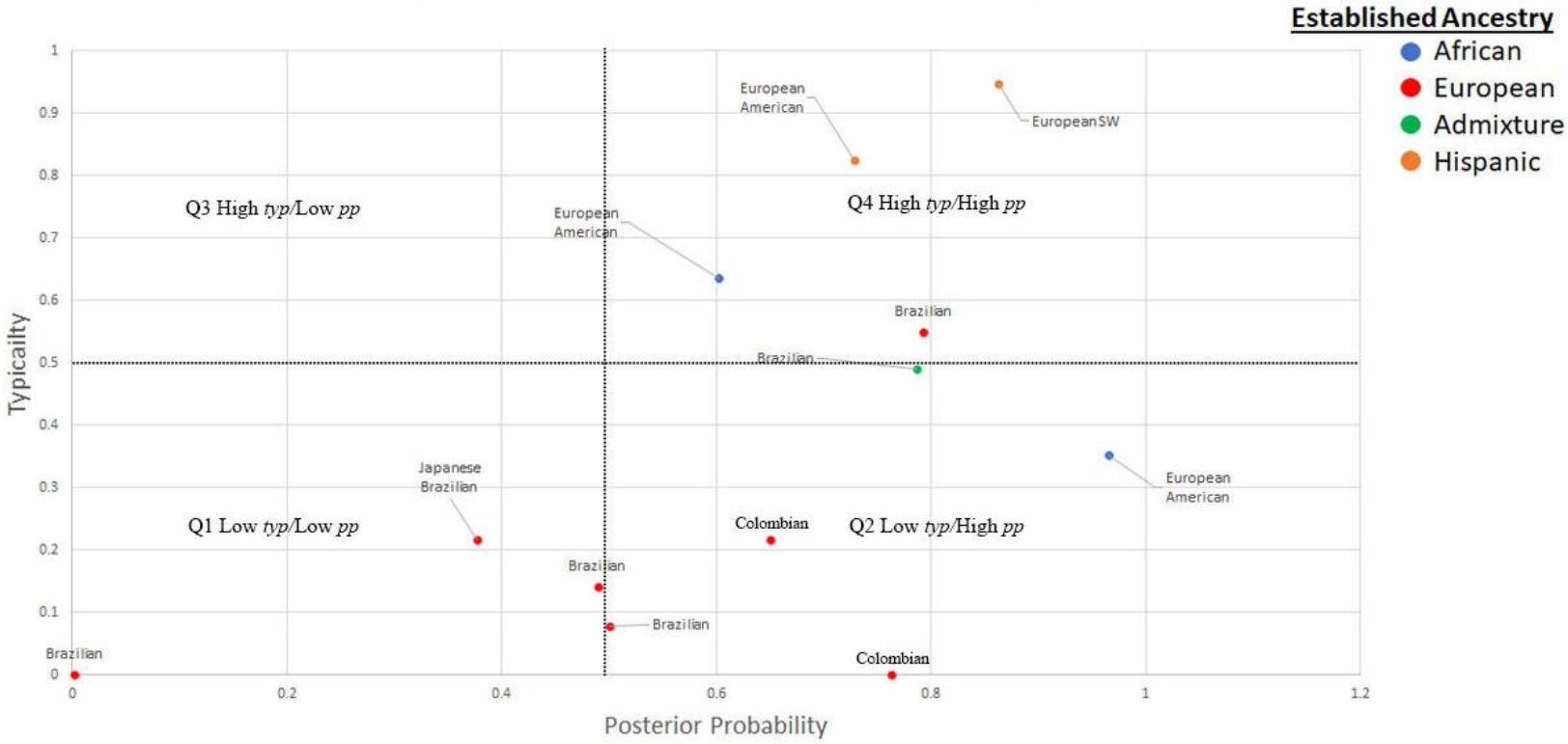

Fig 8. Demonstrates the distribution of incorrect estimated ancestral cases for *pp* and *typ* with 3D-ID Test 2.
The establish ancestry, defined by DNA results is listed as establish ancestry. The written notations identify estimated ancestry by 3D-ID Test 2

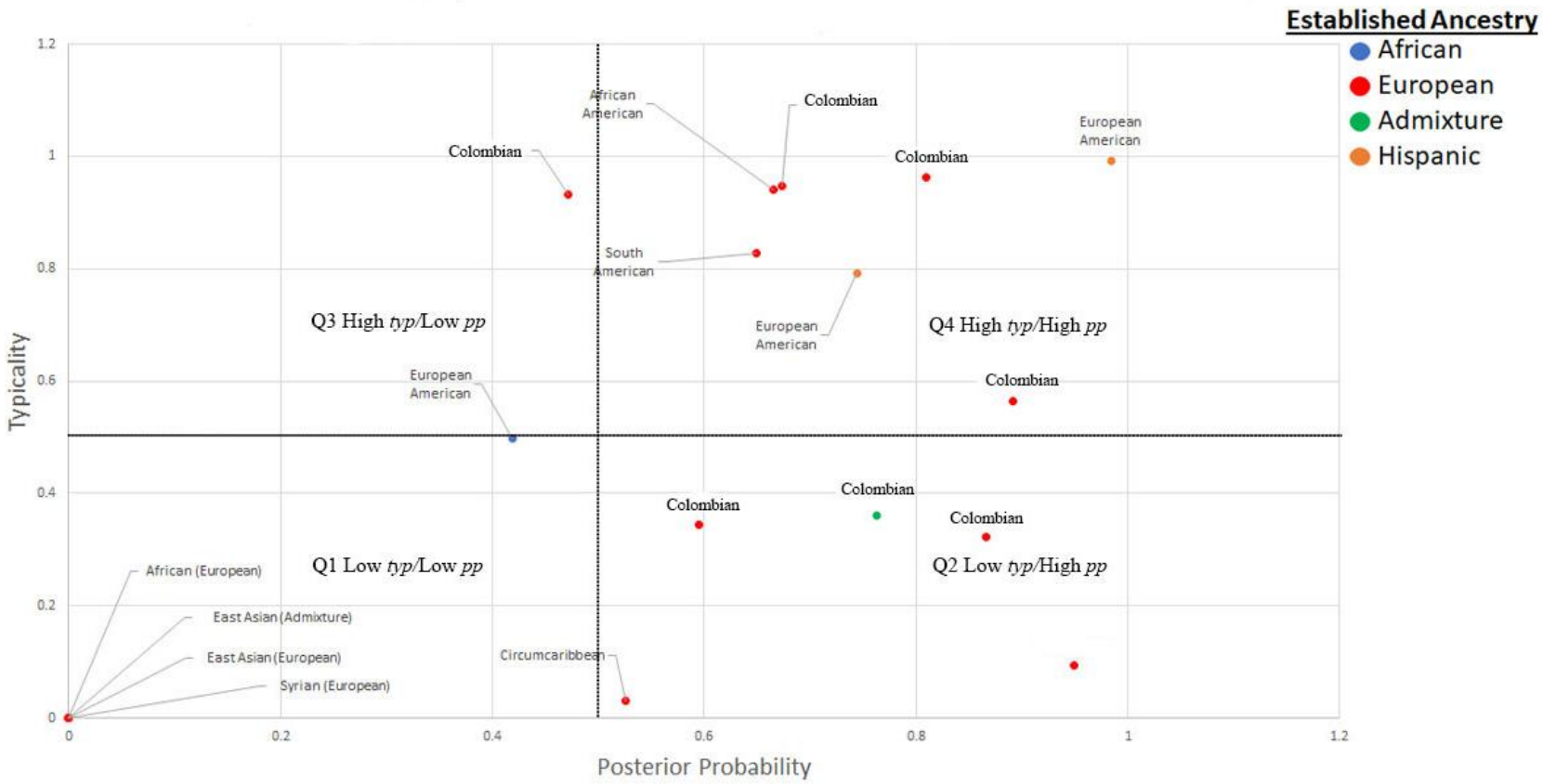

Fig 9. Demonstrates the distribution of incorrect estimated ancestral cases for *pp* and *typ* with 3D-ID Test 3.
The establish ancestry, defined by DNA results is listed as establish ancestry. The written notations identify estimated ancestry by 3D-ID Test 3

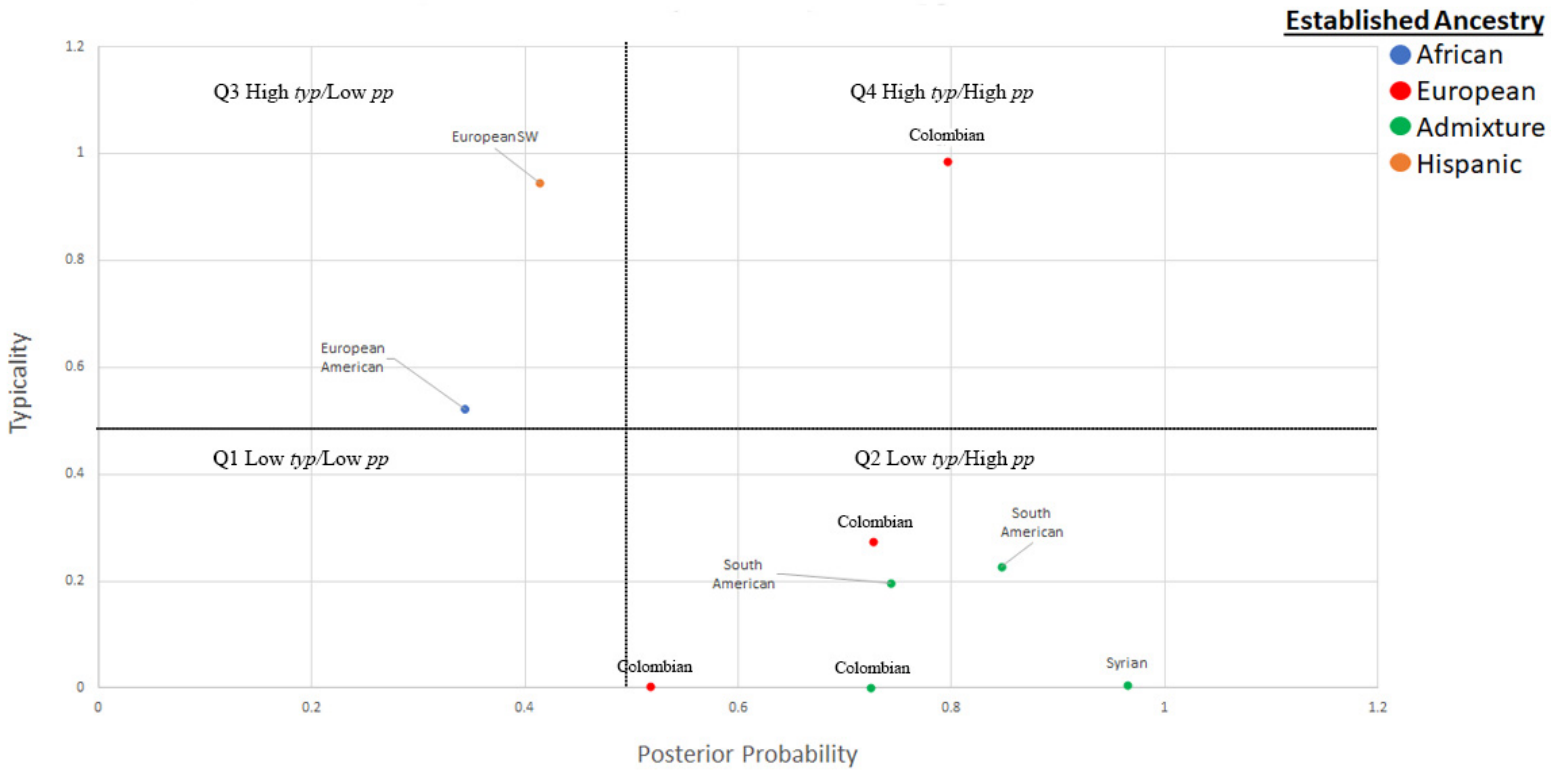

Fig 10. Demonstrates the distribution of incorrect estimated ancestral cases for *pp* and *typ* with 3D-ID Test 4.
The establish ancestry, defined by DNA results is listed as establish ancestry. The written notations identify estimated ancestry by 3D-ID Test 4

Table 9. 3D-ID Correct Classifications by Quadrant

		3D-ID Correct classifications (%)				3D-ID Incorrect classifications (%)			
		Test 1	**Test 2**	**Test 3**	**Test 4**	**Test 1**	**Test 2**	**Test 3**	**Test 4**
	European	12.5	13.0	11.8	30.8	9.1	25.0	16.7	0.0
	African	0.0	0.0	0.0	3.8	0.0	0.0	5.6	0.0
Q1†	Hispanic	0.0	0.0	0.0	0.0	9.1	0.0	0.0	0.0
	Admixture	12.5	8.7	5.9	0.0	0.0	0.0	5.6	0.0
	Total	**25.0**	**21.7**	**17.7**	**34.6**	**18.2**	**25.0**	**27.9**	**0.0**
	European	45.8	39.1	17.6	11.5	18.2	25.0	22.2	22.2
	African	4.2	4.3	5.9	3.8	18.2	8.3	0.0	0.0
Q2†	Hispanic	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Admixture	4.2	8.7	5.9	0.0	9.1	8.3	5.6	44.4
	Total	**54.2**	**52.1**	**29.4**	**15.3**	**45.5**	**41.6**	**27.8**	**66.6**
	European	0.0	4.3	0.0	26.9	18.2	0.0	5.6	0.0
	African	0.0	0.0	0.0	7.8	0.0	0.0	0.0	11.1
Q3†	Hispanic	0.0	0.0	0.0	3.8	0.0	0.0	0.0	11.1
	Admixture	0.0	0.0	0.0	3.8	0.0	0.0	0.0	0.0
	Total	**0.0**	**4.3**	**0.0**	**42.3**	**18.2**	**0.0**	**5.6**	**22.2**
	European	12.5	13.0	29.4	7.8	9.1	8.3	27.8	11.1
	African	8.3	8.7	17.6	0.0	0.0	8.3	0.0	0.0
Q4†	Hispanic	0.0	0.0	0.0	0.0	9.1	16.7	11.1	0.0
	Admixture	0.0	0.0	5.9	0.0	0.0	0.0	0.0	0.0
	Total	**20.8**	**21.7**	**52.9**	**7.8**	**18.2**	**33.3**	**38.9**	**11.1**

†Q1 – *pp*/*typ*,0.5; Q2 – *pp*>0.5, *typ*<0.5; Q3 – *pp*<0.5, *typ*>0.5; Q4 – *pp*/*typ*>0.5

CHAPTER FIVE:

DISCUSSION

When it comes to ancestry estimation in forensic anthropology, there are many biological and technical challenges. Biodistance studies and population affiliation estimated through crania-metric analysis offer the most robust measures. Such a claim is supported in the literature and demonstrated in this study. The application of software programs designed to aid in population estimation through reference samples for comparison are tools for investigators. Nevertheless, the data must be analyzed and interpreted correctly for an accurate result. The choices investigators make for landmark selection, reference populations, and the threshold for acceptable posterior probabilities and typicality can significantly affect the outcome, as demonstrated in this study. The choices' importance is especially true for populations with complex genetic histories as those seen in Central and South America and culturally combined as "Hispanic." Not only is there a lack of adequate reference populations, but the genetic contribution of indigenous peoples, Europeans, and Africans throughout Brazil and South America is also unclear. Likewise, Mayan and indigenous people's genetic influence in Central America and Europeans also makes those reference populations problematic at times, as demonstrated in this study.

Of the eight models evaluated, with different reference populations and thresholds for acceptance and landmarks, the tests generally performed consistently and show that the problem areas continue to be for individuals with combined European and Indigenous genetic histories.

Of all the misclassifications, regardless of the tests, 73 cases fell into this category, where only six of the cases involved African or Asian populations.

Directly comparing Fordisc 3.1 and 3D-ID is not possible due to the different reference populations used in each. Using the 36 solved cases tested in this study, classifications through different Fordisc 3.1 models demonstrated consistent accuracy of just over 80% in each of the tests. However, this was unexpected as Ousley and Jantz (20) point out accuracy should increase when the reference groups selected are reduced, as the computer program has less biodistance to navigate. This reduction did not occur in this current study, and perhaps this cohort had a high affinity with the reference samples from Fordisc 3.1. We examined each reference group further to identify consistencies in correct-classifications/misclassifications to investigate the potentially high affinity to Fordisc 3.1. Accuracy was highest among African populations (100% throughout all tests) and European populations (83%-88%). The individuals with mixed ancestry showed a more significant discrepancy in the accuracy of each test's performance (60%-80%), suggesting one of the parent's genetic contributions to craniometrics form was dominant but not reflected in the classification of the individual. It may also suggest these cases classified as European, based on the craniometrics analysis, and the catch-all category of "Hispanic" is a cultural construct, not a biological one. Over 50% of these cases show there is significant room for error. The results for individuals with "Hispanic" ancestry were by far the weakest, as none of them were correctly classified. The sample size for individuals falling within this category is small; however, the results are consistent with the literature (47,48). This study utilizes real-life cases with the creation of a complete biological profile to include ancestral estimation.

Casework that involves misclassification of individuals as "Japanese" or "Vietnamese" has been an ongoing problem in forensic anthropology (47,48,50). Interestingly, unlike

previously published studies, the Hispanic individuals in this study all classed with European populations (Figure 2-5) and not Asian populations (47,48). The trend for misclassifying individuals from Northern Mexico as European within Fordisc 3.1 is not unheard of (51,52). Understanding common errors are essential because it demonstrates that one can get closer to the correct classification by carefully choosing the right reference populations and landmarks.

The population classifications utilizing 3D-ID, at first, show the overall results appear to be more capricious, with Tests 1 & 2 producing correct classifications at 68.6% and 65.7%, respectively. Test 3 had the lowest accuracy of all the study tests, at 48.6%, whereas Test 4 had the highest accuracy rate at 74.3% (Figure 2). However, this does not tell the whole story, as the models have different functions; we outlined these functions in the methods section. Three of the four tests for 3D-ID accentuated specific reference populations. Did these tests more accurately place cases from associated cohorts at a higher accuracy rate or not?

Model 1 was the shotgun approach and assumed an equal likelihood among all ancestral groups. It is a "blind" attempt to assess population affiliation without utilizing measurements and landmarks highlighting specific reference sample populations. This approach paid off for individuals who had a mixed ancestral background and fell into the "Admix" category, as they classified correctly over 80% of the time. This finding is significant and speaks to the program's strength overall, with these results not achieved using Fordisc 3.1. Consistent with Fordisc 3.1, this model did classify the Hispanic individuals as European American (Figure 6). The Hispanic cohort is affiliated with European populations regularly, suggesting this cohort had a European background, consistent with other individuals living in Central or South America. We found consistent accuracy levels in this model compared to other tests, which we expected for a shotgun approach.

Changing the model to consider specific populations had mixed results; the models utilized the following reference sample populations: Brazilian, MesoAmerican, and European. In Model 2 of 3D-ID, Test 2 used Brazilian and European reference samples by selecting specific landmarks to exploit these reference populations. The overall success rate was 65.7%, with the Hispanic cohort again classifying the members as European. The remaining cases are all classified as either European or Brazilian populations. These results were expected with varied typicalities and posterior probabilities, highlighting the varying confidence 3D-ID demonstrated with each case. In this model, there is an expectation for high performance in the European populations. Although only 69.6% of European cases correctly estimated here, scoring third in the four 3D-ID tests.

3D-ID Test 3 utilized specific landmarks and measurements to focus on MesoAmerican and other reference samples from the Americas. This test misclassified most of the sample, with a correct classification at 48.6%. Only cases in the African cohort produced correct classifications at a high accuracy rate (80%). If Hispanic populations had genuine population affinity, we would expect high success in the Hispanic cohort test; this did not occur as no Hispanic cases were correctly classified.

3D-ID Test 4 is a mixed model pulling from Tests 2 and 3. This model produced a high accuracy rate for the European and African cohorts, with correct classifications of 87% and 80%, respectively. Test 4 had the highest Hispanic accuracy rate (50%). Nevertheless, this test rarely classified the Admix cases correctly (20%). The classification of “Hispanic” individuals was most often with European populations, consistent with all other tests in this study. This test performed well in the European and African cohorts. It also had the most robust performance of

any test in this study for the Hispanic cohort. However, the test struggled with “Admix” individuals.

Throughout all the models tested in 3D-ID and Fordisc 3.1, the European classification consistently had a higher percentage of correct classifications. In Fordisc 3.1 African cohort had a consistent 100% accuracy rate over 10% higher than any of the European cohort. This trend continued in 3D-ID; Admix cohort had an accuracy rate of 10% higher than the European cohort in 3D-ID Test 1 and Test 2. In Test 3, the African cohort outperformed the European cohort by over 30%. Test 4 for 3D-ID highlighted the European cohort’s best performance achieving an 86.5% accuracy rate. There are many correct classifications for the European cohort due to the disproportionate representation of European cases found within the study.

Numerous components are essential in ensuring accuracy when estimating population affinity. Is the reference sample an adequate representation of an actual population built around clines, for example, elevation or distance from the equator, not simple nomenclatures based on continental proximation (Asian, African, European)? Are the typicalities and posterior probabilities high enough to afford confidence for the practitioner running the test? Tables 6 and 9 demonstrate that typicality and posterior probability were not a guarantee of correct population affiliation, but between 70-85% of all correct estimations for all Tests in both 3D-ID and Fordisc fell into Q2 or Q4 (both quadrants with high posterior probability). There was one exception, 3D-ID Test 4 having 22.1% correct estimations in Q2 or Q4.

Ousley and Jantz suggest removing populations with low posterior probability to increase accuracy (20). However, there was no increase in the overall effectiveness in reducing the number of reference populations in this study. In contrast, 3D-ID produced very different results based on population selection. The exclusion and use of specific landmarks to exploit certain

reference sample populations are encouraged by Slice and Ross (27), with the effectiveness illustrated in Figures 8-10. This study's point was to demonstrate how changes in the tests' parameters can affect investigators' outcomes and results on unidentified persons. Thus, the practitioner must be mindful of the model they construct using these programs and utilize the parameters to fit the case's needs best. We recommend using both programs and analyze the results of both for the most accurate estimations in applied casework.

CHAPTER SIX:

CONCLUSION

This study demonstrates that both Fordisc 3.1 and 3D-ID produce accurate population affiliation in over 80% of cases. Changing the model to include selected reference samples, landmarks, measurements and analyzing the typicality and posterior probability results can change the outcome in "ancestry estimation." A holistic approach is often critical in differentiating key components. Analyzing multiple aspects surrounding a case is essential in estimating what data are relevant and skew the results (53).

In the results, there is room for interpretation; these interpretations can be problematic. The programs can estimate the ancestry of an individual. However, the results create an error due to two factors:1) the results are incorrect, but the practitioner accepts them due to high *pp* and *typ*; 2) the practitioner dismisses the results due to low *pp* and *typ*, but the results are correct. As can be seen from Tables 3-4 and Figures 2-9, there are wide-ranging scores within the tests. In many instances, there is a $pp \approx 0$ where the classification is correct and incorrect estimations with $pp \approx 1$; both scenarios can lead to misclassifications.

Forensic Anthropology is an applied science and provides a service to solve real-world questions, assisting in identifying unknown individuals' remains in legal cases, conflicts, natural disasters, and humanitarian efforts. Being confident in population affiliation is critical. While these software programs have shown success in this study with specific populations, there is a need to increase the sample populations used for reference. Also, these populations should be

based on biodistance and not a nomenclature based on US Census designations, such as Hispanic or Asian. There are some conscious efforts to achieve this through databases. One such database is the Forensic Anthropology Case Database or FADAMA, which incorporates real-world cases into the FDB (54,55). The successful application for both programs also rests on continued research and increasing sample sizes and should include increased diversity in the reference populations. Forensic anthropologists must continue to employ these methods as accurately as possible to develop facial approximation. This tool has been significant in the successful identification of the sample used here. Without understanding the genetic history of the unknown individual and what regions of the world may contribute to the skull's shape, it would be impossible to apply any phenotypic data for facial reconstructions.

Nevertheless, this sample of 36 identified individuals did just that. Work still needs to be done to create reference samples within these programs to allow unidentified remains to affiliate with these populations. As populations evolve and adapt, so must our methods within forensic anthropology.

REFERENCES

1. Dunn RR, Spiros MC, Kamnikar KR, Plemons AM, Hefner JT. Ancestry estimation in forensic anthropology: A review. WIREs Forensic Sci 2020;2(4):1–26. https://doi.org/10.1002/wfs2.1369.

2. Maier C, Craig A, Adams DM. Language use in ancestry research and estimation. J Forensic Sci 2021;66(1):11–24. https://doi.org/10.1111/1556-4029.14601.

3. Goodman A. Race Is Real, But It's Not Genetic. Sapiens 2020;:1–10.

4. Spradley MK. Use of craniometric data to facilitate migrant identifications at the United States/Mexico border. Am J Phys Anthropol 2021;(December 2020):1–11. https://doi.org/10.1002/ajpa.24241.

5. Ross AH, Ubelaker DH, Kimmerle EH. Implications of dimorphism, population variation, and secular change in estimating population affinity in the Iberian Peninsula. Forensic Sci Int 2011;206(1–3):214.e1-214.e5. https://doi.org/10.1016/j.forsciint.2011.01.003.

6. Pietrusewsky M. Biological Distance in Bioarchaeology and Human Osteology BT -. In: Smith C, editor. Encyclopedia of Global Archaeology. New York, NY: Springer New York, 2014;889–902.

7. Algee-Hewitt BFB, Kim J, Hughes CE. Thinking computationally about forensics: Anthropological perspectives on advancements in technologies, data, and algorithms. Hum Biol 2018;90(1):5–10. https://doi.org/10.13110/humanbiology.90.1.04.

8. Hughes CE, Juarez CA, Hughes TL, Galloway A, Fowler G, Chacon S. A Simulation for Exploring the Effects of the "Trait List" Method's Subjectivity on Consistency and Accuracy of Ancestry Estimations. J Forensic Sci 2011;56(5):1094–106. https://doi.org/10.1111/j.1556-4029.2011.01875.x.

9. Navega D, Coelho C, Vicente R, Ferreira MT, Wasterlain S, Cunha E. AncesTrees: ancestry estimation with randomized decision trees. Int J Legal Med 2015;129(5):1145–53. https://doi.org/10.1007/s00414-014-1050-9.

10. Edgar HJH. Estimation of Ancestry Using Dental Morphological Characteristics. J Forensic Sci 2013;58(SUPPL. 1):1–6. https://doi.org/10.1111/j.1556-4029.2012.02295.x.

11. Scott G, Navega D, Coelho J, Cunha E, Irish J. rASUDAS: A New Method for Estimating Ancestry from Tooth Crown and Root Morphology. Annu Meet Am Assoc Phys Anthropol 2016;1(1):18–31. https://doi.org/10.13140/RG.2.1.2792.0241.

12. Seguchi N, Dudzik B, Murphy M-M, Schlager S, Archer W, Presnyakova D, et al. 3D Data Acquisition for Bioarchaeology, Forensic Anthropology, and Archaeology. 1st ed. Academic Press, Elsevier, 2019.

13. Brace CL. Region Does Not Mean "Race"—Reality Versus Convention in Forensic Anthropology. J Forensic Sci 1995;40(2):171–5. https://doi.org/10.1520/jfs15336j.

14. Mielke JH, Konigsberg LW, Relethford JH. Comprehending Human Biological Diversity. Human Biological Variation. 2011;3–22.

15. Li Y, Zhao W, Li D, Tao X, Xiong Z, Liu J, et al. EDAR, LYPLAL1, PRDM16, PAX3, DKK1, TNFSF12, CACNA2D3, and SUPT3H gene variants influence facial morphology in a Eurasian population. Hum Genet 2019;138(6):681–9. https://doi.org/10.1007/s00439-019-02023-7.

16. Tise ML, Kimmerle EH, Spradley MK. Craniometric Variation of Diverse Populations in Florida: Indentification Challenges Within A Border State. Ann Anthropol Pract 2014;38(I):111–23. https://doi.org/10.1111/napa.12046.

17. Elliott M, Collard M. FORDISC and the determination of ancestry from cranial measurements. Biol Lett 2009;5(6):849–52. https://doi.org/10.1098/rsbl.2009.0462.

18. Urbanová P, Ross AH, Jurda M, Nogueira MI. Testing the reliability of software tools in sex and ancestry estimation in a multi-ancestral Brazilian sample. Leg Med 2014;16(5):264–73. https://doi.org/10.1016/j.legalmed.2014.06.002.

19. Richard J, Stephen O. Fordisc 3 and Statistical Methods for Estimating Sex and Ancestry. Wiley-Blackwell Companions to Anthropol. A Companion to Forensic Anthropol. 2011.

20. Jantz RL, Ousley SD. Fordisc 3 Computerized Forensic Discriminant Function: Version 3.1. Knoxville, 2005.

21. Slice DE, Ross A. 3D-ID: geometric morphometric classification of crania for forensic scientists. 2009.

22. Ousley S, Jantz R, Freid D. Understanding race and human variation: Why forensic anthropologists are good at identifying race. Am J Phys Anthropol 2009;139(1):68–76. https://doi.org/10.2174/138920312803582960.

23. Austin D, King RE. The Biological Profile of Unidentified Human Remains in a Forensic Context. Acad. Forensic Pathol. 2016;6(3):370–90.

24. Bertsatos A, Christaki A, Chovalopoulou ME. Testing the reliability of 3D-ID software in sex and ancestry estimation with a modern Greek sample. Forensic Sci Int 2019;297:132–7. https://doi.org/10.1016/j.forsciint.2019.02.004.

25. Jurda M, Dvorak M, Urbanova P. Here's looking at you, Fordisc ® …. 2016;(April 2013).

26. Katherine Spradley M, Jantz RL. Ancestry Estimation in Forensic Anthropology: Geometric Morphometric versus Standard and Nonstandard Interlandmark Distances. J Forensic Sci 2016;61(4):892–7. https://doi.org/10.1111/1556-4029.13081.

27. Ross AH, Slice D. 3d-ID home page. NC State Univ. 2018. https://www.iadb.org/en/sector/education/latin-america-and-caribbean-pisa/home.

28. Ta'ala SC. A brief History of the Race Concept in Physical Anthropology. In: Berg GE,

Ta'ala SC, editors. Biological Affinity in Forensic Identification of Human Skeletal Remains. Boca Raton: CRC Press, 2015;17–26.

29. Parsons HR. the Accuracy of the Biological Profile in Casework : an Analysis of Forensic Anthropology Reports in Three Medical Examiners ' Offices. 2017.

30. Gravlee CC, Bernard HR, Leonard WR. Heredity, Environment, and Cranial Form: A Reanalysis of Boas's Immigrant Data. Am Anthropol 2004;105(1):125–38. https://doi.org/10.1525/aa.2003.105.1.125.

31. Caspari R. From Types to Populations: A Century of Race, Physical Anthropology, and the American Anthropological Association. Am Anthropol 2004;105(1):65–76. https://doi.org/10.1525/aa.2003.105.1.65.

32. Morgan LH. Ancient Society. 2nd ed. Cleveland: Meridian, 1967.

33. Montagu A. The concept of race. Am Anthropol 1962;64(1):919–28. https://doi.org/10.1177/039219216701505906.

34. Darwin CR. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. London : John Murray, 1859, 1859.

35. Livingstone FB. On the Non-Existence of Human Races. Curr Anthropol 1962;3(3):279–81.

36. Greaves M. Was skin cancer a selective force for black pigmentation in early hominin evolution? Proc R Soc B Biol Sci 2014;281(1781). https://doi.org/10.1098/rspb.2013.2955.

37. Smay D, Armelagos G. Galileo Wept: A Critical Assessment of the Use of Race in Forensic Anthropology. Transform Anthropol 2005;9(2):19–29. https://doi.org/10.1525/tran.2000.9.2.19.

38. Rhine S. Skeletal Criteria for Racial Attribution. NAPA Bull 2005;13(1):54–67. https://doi.org/10.1525/napa.1993.13.1.54.

39. Konigsberg LW, Herrmann NP, Wescott DJ, Kimmerle EH. Estimation and evidence in forensic anthropology: Age-at-death. J Forensic Sci 2008;53(3):541–57. https://doi.org/10.1111/j.1556-4029.2008.00710.x.

40. Maiers J. Revisiting the Concepts of Race. AAFS. Anaheim, 2020.

41. Bethard JD, DiGangi EA. Letter to the Editor—Moving Beyond a Lost Cause: Forensic Anthropology and Ancestry Estimates in the United States. J. Forensic Sci. 2020;65(5):1791–2.

42. Stull KE, Bartelink EJ, Klales AR, Berg GE. Commentary on: Letter to the Editor—Moving Beyond a Lost Cause: Forensic Anthropology and Ancestry Estimates in the United States. J Forensic Sci 2020;65(5):1791–2. https://doi.org/10.1111/1556-4029.14513.

43. Ross AH, Pilloud MA. Blinded by the White. Adulteration of Ancestry. A time for reform. 2021.

44. Kimmerle EH, Tise ML, Humphries AL. Data Collection Protocol For Human Identification. Tampa Fl, 2012.

45. Chakraborty R, Weiss KM. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc Natl Acad Sci U S A 1988;85(23):9119–23. https://doi.org/10.1073/pnas.85.23.9119.

46. Kamenov GD, Kimmerle EH, Curtis JH, Norris D. Georeferencing a Cold case victim with lead, strontium, carbon, and oxygen. Ann Anthropol Pract 2014;(December):137–54.

47. Dudzik B, Jantz RL. Misclassifications of Hispanics Using Fordisc 3.1: Comparing Cranial Morphology in Asian and Hispanic Populations. J Forensic Sci 2016;61(5):1311–8. https://doi.org/10.1111/1556-4029.13123.

48. Dudzik B. Examining Cranial Morphology of Asian and Hispanic Populations Using Geometric Morphometrics for Ancestry Estimation. Forensic Anthropol 2019;2(4):304–15. https://doi.org/10.5744/fa.2019.1022.

49. Konigsberg LW, Frankenberg SR. Typicality and predictive distributions in discriminant function analysis. Hum Biol 2018;90(1):31–44. https://doi.org/10.13110/humanbiology.90.1.01.

50. Go M, Jones A, Algee-Hewitt B, Dudzik B, Hughes C. Classification Trends among Contemporary Filipino Crania Using Fordisc 3.1. Forensic Anthropol 2019;2(4):293–303. https://doi.org/10.5744/fa.2019.1005.

51. Hughes CE, Dudzik B, Algee-Hewitt BFB, Jones A, Anderson BE. Understanding (Mis)classification Trends of Latin Americans in Fordisc 3.1: Incorporating Cranial Morphology, Microgeographic Origin, and Admixture Proportions for Interpretation. J Forensic Sci 2019;64(2):353–66. https://doi.org/10.1111/1556-4029.13893.

52. Hughes CE, Tise ML, Trammell LH, Anderson BE. Cranial morphological variation among contemporary Mexicans: Regional trends, ancestral affinities, and genetic comparisons. Am J Phys Anthropol 2013;151(4):506–17. https://doi.org/10.1002/ajpa.22288.

53. Hughes CE, Algee-Hewitt BFB, Reineke R, Clausing E, Anderson BE. Temporal Patterns of Mexican Migrant Genetic Ancestry: Implications for Identification. Am Anthropol 2017;119(2):193–208. https://doi.org/10.1111/aman.12845.

54. Juarez CA, Hughes CE, Yim A Di. Technical note: A report on the Forensic Anthropology Database for Assessing Methods Accuracy. Am J Phys Anthropol 2021;174(1):149–50. https://doi.org/10.1002/ajpa.24167.

55. Hughes CE, Juarez CA. Learning from Our Casework: Forensic Anthropology Case Database for Assessing Methods Accuracy (FADAMA). NIJ 2018-DU-BX-0213. 2018. https://www-app.igb.illinois.edu/sofadb/faq.php#scope (accessed February 27, 2021).