11-10-2021

Comparing Genetic Variation among Latin American Immigrants: Implications for Forensic Casework in the Arizona- and Texas-México Borderlands

Briana T. New
SNA International

Bridget F.B. Algee-Hewitt
Stanford University

M. Katherine Spradley
Texas State University, San Marcos

Lars Fehren-Schmitz
University of California, Santa Cruz

Cris C. Hughes
University of Illinois Urbana–Champaign

See next page for additional authors
Follow this and additional works at: https://digitalcommons.wayne.edu/humbiol_preprints

Authors

Briana T. New, Bridget F.B. Algee-Hewitt, M. Katherine Spradley, Lars Fehren-Schmitz, Cris C. Hughes, Bruce E. Anderson, Marek E. Jasinski, Joanna Arciszewska, Grażyna Zielińska, Maria Szargut, Sandra Cytacka, and Andrzej Ossowski

Comparing Genetic Variation among Latin American Immigrants: Implications for Forensic Casework in the Arizona- and Texas-México Borderlands

Briana T. New,[1]* Bridget F.B. Algee-Hewitt,[2] M. Katherine Spradley,[3] Lars Fehren-Schmitz,[4,5] Cris E. Hughes,[6] Bruce E. Anderson,[7] Marek E. Jasinski,[8,9] Joanna Arciszewska,[10] Grażyna Zielińska,[10] Maria Szargut,[10] Sandra Cytacka,[10] and Andrzej Ossowski[10]

[1]SNA International supporting the Defense POW/MIA Accounting Agency, Joint Base Pearl Harbor, Hickam, Hawaii, USA.

[2]Humanities and Sciences Interdepartmental Programs, Stanford University, Stanford, California, USA.

[3]Anthropology Department, Texas State University, San Marcos, Texas, USA.

[4]UCSC Genomics Institute, University of California, Santa Cruz, Santa Cruz, California, USA.

[5]Anthropology Department, University of California, Santa Cruz, Santa Cruz, California, USA.

[6]Anthropology Department, University of Illinois Urbana–Champaign, Urbana, Illinois, USA.

[7]Pima County Office of the Medical Examiner, Tucson, Arizona, USA.

[8]Department of Historical and Classical Studies, NTNU Norwegian University of Science and Technology, Trondheim, Norway.

[9]Falstad Centre, Ekne, Norway.

[10]Department of Forensic Genetics, Pomeranian Medical University in Szczecin, Szczecin, Poland.

*Correspondence to: Briana T. New, SNA International supporting the Defense POW/MIA Accounting Agency, 590 Moffet St., Joint Base Pearl Harbor, Hickam, HI 96853 USA. E-mail: btrnew@gmail.com.

Short Title: Genetic Variation among Latin American Immigrants

KEY WORDS: TEXAS-MÉXICO IMMIGRANTS, ARIZONA- MÉXICO IMMIGRANTS, POPULATION STRUCTURE, GENETIC VARIATION, FORENSIC IDENTIFICATION, REGION OF ORIGIN.

Abstract

The humanitarian crisis on the United States-México border is a long standing and evolving crisis in which nearly 8,000 deaths have been reported in the last two decades. These deaths are largely distributed across the Arizona-México and Texas-México border regions where demographic trends for immigrants attempting to cross into the U.S. have shifted dramatically. The demographic change and volume of immigrants seeking shelter in the U.S. presents new challenges for the forensic practitioners entrusted with the identification of individuals who lose their lives during the final segment of their journey. Within this Border context, the present study investigates how genetic variation inferred from forensically significant microsatellites can provide valuable information on regions of origin for unidentified remains on the group level. To explore how we can mobilize these genetic data to inform identification strategies, we conduct a comparative genetic analysis of identified and unidentified immigrant cases from the Arizona- and Texas-México contexts, as well as 27 other Latin American groups. Allele frequencies were utilized to calculate FST, and relationships were visually depicted in a multidimensional scaling plot. A Spearman correlation coefficient analysis assessed the strength and significance of population relationships and an agglomerative clustering analysis assessed population clusters. Results indicate that Arizona-México immigrants have the strongest relationship (>80%) with groups from El Salvador, Guatemala, México, and an indigenous group from Southern México. Texas-México immigrants have the strongest relationships (>80%) with groups from Belize, Colombia, Costa Rica, El Salvador, Guatemala, Honduras, and Nicaragua. These findings agree with, and are discussed in comparison to, previously reported demographic trends, population genetics research, and population history analyses. We emphasize the utility and necessity of

coupling genetic variation research with a nuanced anthropological perspective for identification

processes in the U.S-México border context.

The humanitarian crisis on the shared border between the United States and México has spanned decades, claiming the lives of at least 7,805 people between 1998 and 2019 (United States Border Patrol 2019a). This accounting is likely a vast underestimation on the scope of this crisis because for remains to be appropriately counted they must be, first, discovered across expansive stretches of public and private land, next, recognized as a person, most often a refugee, from Latin America seeking to cross the border into the United States, and, last, documented in consistent, centralized systems (Anderson 2008; Anderson and Parks 2008; Gocha et al. 2018; Martinez et al. 2013). Moreover, simple tabulation of immigrant deaths by border region provides little context for understanding the demographic distributions, such as region of origin, for those peoples crossing and dying along the United States' southern border (Algee-Hewitt et al. 2018; Anderson 2008; Anderson and Parks 2008; Hughes et al. 2017).

While there are region specific efforts to document mortality over time and provide case recovery details that may aid in identification of Latinx immigrants in Arizona (Humane Borders 2020), thus far, the United States Border Patrol (USBP) is the only official source of information for immigrant "illegal alien" apprehensions by sector, apprehensions by citizenship, and summary statistics for deaths by border region (U.S. Customs and Border Protection 2020). This representation is problematic because the USBP do not report their methodology, share the origin of their data, and have had inconsistencies in their border death tabulations when compared to that of a medical examiner's office (Reineke and Halstead 2017). Furthermore, the border patrol is not responsible for identifying the remains of an immigrant decedent, rather it is the responsibility of forensic identification practitioners, operating in vastly different jurisdictions, whose systems are regionally specific (Gocha et al. 2018; Reineke and Halstead 2017).

To facilitate the identification of an immigrant decedent, forensic anthropologists must develop a biological profile against which a missing persons report can be compared. These evaluations include well established methods for estimating personal identity parameters like sex, stature, and age-at-death. In the context of the Border crisis, where individuals travel from many different countries, the parameter of ancestry is arguably the most critical aspect of the biological profile for unidentified immigrant remains, yet it is at the same time the most elusive. The challenge lies in the fact that the often used category of Hispanic is difficult to support with classification statistics and even more difficult to reconcile with what is known about the complex histories of Latin America and the range of biological signatures that such variation produces (Algee-Hewitt 2017; Algee-Hewitt et al. 2018; Dudzik and Jantz 2016; Hughes et al. 2019; Tise et al. 2014; Spradley 2014; Spradley 2016). Accordingly, ancestry, as defined for forensic anthropology, is understood best in microgeographic terms – producing a "biogeographic profile" that emphasizes "place of recent origin" or "sending region" (Algee-Hewitt et al. 2020).

To obtain accurate and precise information on region of origin, forensic practitioners can attempt to utilize personal effects (i.e. ID cards, currency, personal notes, religious icons, etc.) or adopt a life history approach (Anderson 2008; Birkby et al. 2008; Soler et al. 2019; Spradley et al. 2019). These approaches are significant because they centralize the personal, cultural, and social identities of the individual. In adopting concepts of postmortem human dignity for the identification of immigrants, forensic practitioners can understand more about an individual, as perceived by themselves, in what they choose to carry with them (Spradley et al. 2019). Unfortunately, these items are not always present, or reliable, and the biocultural signals of lived

experience are not always legible (Anderson 2008; Birkby et al. 2008; Soler et al. 2019; Spradley et al. 2019).

In order to estimate ancestry information for a given Border case, researchers have developed methods that utilize craniometric, morphoscopic (discrete) trait, and dental variation using, primarily, skeletal remains from México, Guatemala, and U.S. Hispanic groups to differentiate among populations (Algee-Hewitt et al. 2020; Hefner et al. 2015; Hughes et al. 2013; Maier and George 2020; New 2018; Spradley 2014; Spradley 2016). Currently, however, comprehensive Latin American skeletal references that are reflective of the shifting trends in sending regions are difficult to obtain and, therefore, limit comparative assessments of skeletal morphology. These constraints demand more innovative solutions for quantifying Latinx variation broadly and distilling these patterns down to the level at which something useful for identification can be said about the individual. To this point, there is promising new work emerging. Machine learning models are being developed that merge relative estimates of triparental ancestry, case year, and geospacial data for the location of recovery to infer place of origin and subsequently map immigration pathways from home region to the Arizona border (Algee-Hewitt et al. 2020); while research focusing on the utility of stable isotopes in the context of unidentified remains is demonstrating the potential for excluding geographic regions when isotopic values are inconsistent with the remains in question (Ammer et al. 2020; Bartelink et al. 2020; Kramer et al. 2020). However, these isotopic methods are still developing and have yet to be comprehensively compared to individuals of known origin with consideration for water stress and resource globalization (Juarez et al. 2020), and the biogeographic models require the integration of data representing cases across multiple Border states.

Developing methodologies that provide access to information on country of origin is therefore critical, especially when placed within the context of identification strategies. If a family is not readily located within the United States, to which countries do we send our reports? To which communities do we ask non-governmental organizations (NGOs) to provide our case information? Where should we seek more missing persons reports and family reference samples?

To help resolve the challenge of deducing region of origin for unidentified immigrant remains, there are many good reasons to study the genetic variation among Latin American populations as inferred by microsatellites (short tandem repeats, or STRs) that are routinely typed for forensic casework. First, there is a great amount of genetic diversity that is geographically structured (Algee-Hewitt et al. 2018; Algee-Hewitt 2018; Hughes et al. 2017; Moreno-Estrada et al. 2014; Rangel-Villalobos et al. 2016; Rubi-Castellanos et al. 2009; Salazar-Flores et al. 2015; Wang et al. 2008). Second, there is a plethora of population data available for comparison (relative to skeletal samples) that have been generated over the last two decades which offers the opportunity for more comprehensive sampling, as is necessary for studying Latin American population variation. Finally, recent work has corroborated the concordance between genetic and skeletal estimates of place of origin; moreover, these forensic STRs, while chosen for their power to make individual identifications, are also valuable for the ancestry information that they convey (Algee-Hewitt 2016; Algee-Hewitt et al. 2016; Hughes et al. 2013; New 2018).

In this paper, we ask: How can studying contemporary Latin American genetic diversity, using STR data, help us better determine the various origins of U.S.-México border immigrants? In doing so, this paper highlights how population genetics research into the geographic structure observable in forensic STRs can play a fundamental role in the case investigation process and, in

turn, a key role in enabling positive identifications when applied in the Border context. By combining the complementary expertise of forensic anthropology and forensic genetics, we can mobilize genetic data and methods in a way that guides and supports the efforts of forensic practitioners and the organizations with which they collaborate.

Comparing Two Critical Border Contexts

In this paper, we focus our sampling on two major epicenters with distinct waves of migration from the late 1990s to the present: the Tucson region of the Arizona-México border and the Rio Grande Valley region of the Texas-México border. These waves of migration, and associated deaths, are largely represented by two investigative organizations.

First, we draw on data provided by the Pima County Office of the Medical Examiner (PCOME) in Tucson, Arizona. The PCOME is a county supported agency that is responsible for investigating the majority of immigrant deaths discovered in the Tucson border region (Anderson and Spradley 2016; Gocha et al. 2018). As of 2019, the PCOME has examined over 2,000 individuals known to be immigrants and over 1,000 cases suspected to be the remains of immigrants who died in southern Arizona after entering the United States (Pima County Office of the Medical Examiner 2019). Of these cases, the PCOME report that, from 2000–2019, 81% of identifications were comprised of Mexicans, 11% Guatemalan, 3% Salvadoran, 3% Honduran, and the remaining ~2% distributed across various other populations (Pima County Office of the Medical Examiner 2019).

The second organization we draw on is Operation Identification (OpID). OpID, founded in 2013, is a non-governmentally affiliated organization that operations within the Forensic Anthropology Center at Texas State University and is focused on identifying the remains of

presumed migrants discovered in South Texas (Anderson and Spradley 2016; Gocha et al. 2018; Spradley and Gocha 2020). Because the majority of counties along the Texas border operate under a county specific justice of the peace system, rather than the centralized medical examiner system of Arizona, forensic anthropologists working with OpID are involved with locating, exhuming, and providing identification efforts for the long-term dead in South Texas near the Rio Grande Valley border (Gocha et al. 2018; Spradley and Gocha 2020). OpID faces numerous obstacles in the identification process, one of which has been substantial variation in the regions of origin for identified individuals when compared to PCOME. Of the 311 cases supervised by OpID since 2013, 38 individuals have been identified as of July 2019. Of those 38 individuals, 32% (n=12) were Guatemalan, 29% were Mexican (n=11), 26% (n=10) were Salvadoran, 8% were Honduran (n=3), 3% were Ecuadoran (n=1), and 3% were (n=1) Nicaraguan.

Death and Apprehensions in Light of Immigration Policies

Increase in immigrant deaths were driven first by a series of anti-immigration policies enacted by the United States to actively deter attempted crossings (Cornelius 2001; Reineke and Halstead 2017; Soto and Martínez 2018). The effects of these anti-immigration policies redirected immigrants away from popular, arguably safer, crossing points in California and Texas and instead funneled people into the most treacherous, deadly terrain of Arizona and, more recently, Texas. Rather than reducing rates of attempted immigration, the deterrent measures of the "funnel effect" in the short period between 1990–2005 resulted in a 20% increase in the number of immigrant death cases investigated by PCOME in Tucson, Arizona (Rubio-Goldsmith et al. 2006). These rates continued increasing until they peaked at 251 reported deaths in 2010, nearly double that of any other border region in any year prior (United States Border Patrol 2019a;

Martinez et al. 2013). However, in 2012 the highest number of deaths shifted from Arizona to Texas, where reported deaths for all Texas regions reached 277 deaths, signaling the beginning of a new migration trend (United States Border Patrol 2019a). For 2019, deaths for all Texas regions continue to surpass the Tucson sector of the border with USBP reporting 225 deaths and the PCOME reporting 153 deaths (Pima County Office of the Medical Examiner 2019; United States Border Patrol 2019a).

Along with shifts in deaths, the new wave of migration brought a shift in the demographic makeup of immigrants. The PCOME estimates that of the 2,238 remains recovered in their jurisdiction between 1990–2012, 80% of the identified were male, 53% of whom were between the ages of 20–39, and 82% of whom originated from México (Martinez et al. 2013). However, in 2014, the total number of "non-Mexicans" apprehended by the USBP surpassed apprehensions of individuals from México (United States Border Patrol 2019b). It is believed that this increase in non-Mexican national apprehensions is largely driven by families and unaccompanied children, rather than lone men, fleeing violence in the Northern Triangle (Guatemala, Honduras, El Salvador) who are seeking asylum within the United States (Soto and Martínez 2018).

To further illustrate the significant shift in immigrant regions of origin, we compared USBP apprehension numbers by citizenship between the Tucson and Rio Grande Valley sectors of the border for 2012 and 2019 (United States Border Patrol 2019b). In 2012, the greatest number of apprehensions came from individuals of Mexican origin at the Tucson sector (n=102,303), followed by individuals of Mexican origin at the Rio Grande Valley sector (n=47,823) (Figure 1). Individuals from all other countries were apprehended in much lower proportions.

In 2019, however, apprehensions of individuals from México dropped to about 29,000 for both sectors (Figure 1). The other most represented countries (Honduras, Guatemala, El Salvador, and Nicaragua) increased exponentially in Rio Grande Valley. Apprehensions of Hondurans reached over 146,000, while apprehensions of Guatemalans were over 81,000, Salvadorans were over 59,000 and Nicaraguans were over 8,000. This information indicates that, demographically, immigrants crossing through the Arizona-México and Texas-México borderlands represent two waves of migration with different sending regions.

Genetics Research for Case Investigation and Identification

By quantifying forensic genetic and craniometric traits in Border death cases from PCOME, research on the patterning of this data finds that the genetic and skeletal variation observed for immigrants in the Arizona context is structured temporally, geographically, and demographically (Algee-Hewitt 2016; Algee-Hewitt et al. 2018; Hughes et al. 2013; Hughes et al. 2017). These data are patterned temporally such that proportions of European vs. Indigenous genetic admixture have shifted over time. They are patterned geographically such that genetic variation and craniometric variation are organized in a North-to-South cline, with individuals from geographically proximate regions of México sharing stronger population relationships and similar admixture proportions. And, they are structured demographically such that the number of immigrants from the Central and Southern regions of México has increased over time. In each of these instances, utilizing forensic genetic markers, or craniometric proxies, proved invaluable for illuminating bio-social factors modulating identification trends and for assessing how, in this Arizona context, immigrant morphogenetic data are geographically structured.

Despite the similarly long history of crossings along the Texas Border, the equally critical need for making identifications from the growing number of deceased in this region, and the just as difficult challenges that changing demographics pose to forensic anthropological efforts in repatriation, only preliminary analyses of genetic data for OpID Border cases have been presented thus far (New et al. 2019; Spradley et al. 2019). The present paper represents the first published study to both investigate the genetic variation in the Texas- and Arizona-México border contexts simultaneously and draw comparisons between the populational make-up of the Arizona and Texas Border deaths.

We hypothesize that investigating the genetic variation among immigrant remains will reveal information on genetic structure within these arguably unique deceased groups of individuals that can assist in the case investigation as well as in the populational and personal identification processes. Using STR loci that are routinely typed for forensic identification, we will compare the genetic data of identified and unidentified immigrants from the Arizona-México border region (PCOME) and the Texas-México border region (OpID) to preexisting data obtained for a wide sampling of Latin American populations. In doing so, we reveal the similarities and differences in immigration trends between PCOME and OpID from a genetics perspective, highlight the accuracy of STR genetic data for assessing population relationships, and demonstrate the potential for using genetic variation as an important resource for accessing information on region of origin for unidentified remains. We argue that understanding immigrant variation and its relationship to geographic location will allow anthropologists, and their collaborators, to tailor their identification efforts to the most appropriate regions and develop more efficient investigative strategies.

Materials

Forensic DNA Markers

This study analyzes 15 STR genetic markers obtained from bone samples of identified and unidentified U.S.-México immigrant decedents as well as other Latin American groups from throughout México and Central/South America (Table 1). The 15 STR genetic markers were selected for their standardized use in the United States' Combined DNA Index System (CODIS) for forensic identification (Butler and Hill 2012). The markers are CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, TH01, TPOX, vWA, D2S1338, and D19S433. While there has been some debate on the most appropriate and informative genetic markers for population structure inference (Liu et al. 2005; Rosenberg et al. 2003), recent research has shown that highly polymorphic microsatellites are stronger in their population structure detection despite, or perhaps because of, their individuating characteristics (Algee-Hewitt et al. 2016; Haasl and Payseur 2011).

Cases by Region

The Arizona-México immigrant sample (designated PCOME) derives from unidentified and identified individuals whose remains were recovered between 1972 and 2013 (Hughes et al. 2017). The identified individuals included in this analysis are exclusively from regions of México, including Northwest México (n=32), Central México (n=45), and Southeast México (n=27). The remaining individuals are unidentified with unknown regions of origin (n=238).

The Texas-México immigrant sample (designated OpID) includes genetic data from unidentified and identified individuals whose remains were recovered between 2013 and 2017. The sample consists of 61 individuals. Most of the sample is currently unidentified with

unknown regions of origin (n=45). The remaining individuals are from Ecuador (n=1), El Salvador (n=6), Guatemala (n=1), Honduras (n=2), México (n=5), or Nicaragua (n=1). The number of identified individuals is low therefore they have been included in the overall sample for OpID.

Each additional sample included in the following analyses were selected because of their expected shared population structure with the Texas-México and Arizona-México immigrants, as well as for the availability of all 15 STRs (Table 1). These samples include previously generated STR datasets from Belize, Colombia, Costa Rica, El Salvador, Guatemala, Honduras, various regions of México, Nicaragua, and Venezuela (Flores et al. 2015; Porras et al. 2008; Rodríguez et al 2007; Morales et al. 2004; Martinez-Espin et al. 2006; Matamoros et al. 2008; Barrot et al. 2005; González-Martín et al. 2008; Gorostiza et al. 2007; Juárez-Cedillo et al. 2008; Luna-Vazquez et al. 2005; Rangel-Villalobos et al. 2013; Sánchez et al. 2005; Locia-Aguilar et al. 2018; Nuñez et al. 2010; Bernal et al. 2006).

Terminology

To assess the patterns of ancestry for both the Texas-México and Arizona-México immigrants, each sample was labeled based on their apparent quantity of European admixture, using Indigenous as the baseline for this relative comparison. Therefore, we designate samples that have reported or presumed low European admixture as Indigenous; and we refer to samples that were otherwise labeled by the term Mestizo, as persons with High European Admixture (HEA). We utilize the descriptor of HEA to reject any engagement with the term Mestizo, which is a vestige of the *casta* system that emerged along with the Spanish Empire's occupation of Latin America (Gutiérrez 2015). "Mestizo" is, therefore, problematic for the colonial history and racial

or ethnic hierarchy that it represents, the system of inequality that its use continues to perpetuate, and its ill-defined relationship to ancestry estimates reported in the literature. In the context of post-colonial contact México, Mestizo was initially defined as Mexican ancestry for at least 3 generations with a Spanish-derived last name (Rubi-Castellanos et al. 2009; Salazar-Flores et al. 2015; Sánchez-Serrano 1996). In genetic literature it is most often used to refer generally to Latin American groups with greater European admixture under the tripartite Indigenous/African/European admixture model (Rubi-Castellanos et al. 2009; Wang et al. 2008;). While most of the studies utilized in this analysis define whether groups in their analyses are of Indigenous origin or are "Mestizo," many researchers provide little explanation of the meaning or source of their terminology. It is especially unclear whether these group designations are prescribed by the researchers or self-reported by individuals comprising the sample. Here we use "ancestry" to refer only to the reported ancestral group designations that are based on expected Indigenous or European parental ancestry proportions, as inferred from computational analysis. These designations may or may not overlap with how individuals within the study populations were identified or would identify themselves.

Methods

Pairwise genetic distances (F_{ST}) were estimated among all populations using published or calculated allele frequencies in GenAlEx 6.503 (Peakall and Smouse 2012). F_{ST} remains the standard measure for interpopulation analyses though other measures, such as G_{ST}, R_{ST}, or D, have been identified as potential alternatives (Ma et al. 2015; Meirmans and Hedrick 2011). However, the relatively low F_{ST} values classically associated with highly polymorphic markers, such as the CODIS STRs utilized here, should not be interpreted to indicate a lack of power for

inferring population relationships or estimating ancestry components (Algee-Hewitt et al. 2016).

Indeed, work targeting this issue specifically has shown that markers used for individuation in the forensic context also enable the discovery of population structure and convey information on ancestry (Algee-Hewitt et al. 2016).

The genetic distances were visually represented in a multidimensional scaling (MDS) plot and a Kruskal's stress test was used to assess the fit of the data to the plot. A Spearman correlation coefficient analysis of the genetic distances measured the strength and significance of the intrapopulation relationships (alpha = 0.01). The analysis produces correlations, levels of significance for each population relationship, and coefficients of determination to measure the proportion of the variance that is predictable from each variable. Groups that differed significantly from the majority of other groups in the analysis were removed because, as Moreno-Estrada et al. (2014) also found, the strength of the genetic variation for the Indigenous groups, as produced by genetic drift, masks the variation within HEA groups. An agglomerative hierarchical clustering analysis was implemented on this reduced sample to evaluate population clusters. All of these analyses were conducted using XLSTAT 2020.1.3 (Addinsoft 2020).

Results

Pairwise genetic distances for all Latin American groups are visualized in Figure 2. The Kruskal's stress value for the MDS is 0.133, which indicates good fit of the data to the visualization. In the MDS, Dimensions 1 and 2 are largely driven by the strength of the genetic variation for Indigenous groups while the HEA groups demonstrate a more linear relationship across dimensions. The plot highlights that most Indigenous groups within this analysis differ greatly from the immigrant or HEA groups. Furthermore, we underscore that Southeast Mexican

individuals from the PCOME sample are most distant from the higher European admixture groups and the other identified immigrant samples. All other immigrant groups cluster most closely to other HEA samples and one Indigenous sample (Choles) from southern México. Generally, within the HEA cluster, the Mexican HEA groups cluster tightly together with the Choles, PCOME, El Salvador, and Guatemala while OpID and the other Central/South American groups cluster more closely together.

The Spearman correlation coefficient analysis found that all populations, except for the majority of Indigenous groups, have moderate to high ($0.58 < rho < 0.99$) correlations and statistically significant associations ($p < 0.0001$). Due to lack of statistically significant associations, the majority of Indigenous groups (Otomi SM and XM, Huasteco, Tarahumara, Tepehuano, and Mexicaneros) were removed from any additional analyses. To assess how much of the variation can be explained by the pairwise relationships in this analysis as well as the strength of the predicted relationships, we address only the coefficients of determination (R^2) in greater detail and limit our reporting to groups with $R^2 < 80\%$ to either OpID or PCOME (Table 2). This 80% cutoff was established to highlight the clear differentiation of stronger genetic relationships above 80% from the other population relationships in the analysis. Our results indicate that the OpID sample is most closely related to the Belize, Colombia (Caldas), Costa Rica, El Salvador, Guatemala, Honduras, and Nicaragua samples ($R^2 < 80\%$). Between the OpID and the PCOME identified and unidentified samples, no genetic relationship exceeds 68%. The PCOME identified from Central México, Northwest México, and unidentified individuals share the strongest genetic relationships with each other and HEA samples from México and El Salvador. Only the PCOME identified from Central México and unidentified samples share a genetic relationship greater than 80% with Guatemala. Finally, and as above, the PCOME

identified sample from Southeast México does not share a strong relationship with all other groups in the analyses except for the PCOME unidentified and Indigenous Choles sample from South México. Aside from Guatemala and El Salvador, the Central American groups and the Colombian sample have an inverse relationship, where all relationships fall below 68%, with all groups from México.

To further demonstrate the clustering of the HEA groups, outlier Indigenous samples were removed and an agglomerative hierarchical clustering analysis produced a dendrogram with 6 classes (Figure 3). Indigenous groups from México (Huichol & Cora, Tepehua, Mayos) form their own classes. However, of those Indigenous groups, the Mayos demonstrate the least genetic difference from the HEA groups. The fourth class includes the OpID Texas-México immigrants and consists of all non-Mexican Latin American groups included in this analysis. The PCOME identified cases from Southeast México appears independent with its own class. The final class includes the PCOME unidentified and identified cases from Northwest/Central México, Mexican HEA groups, an Indigenous group from Central México (Choles). Overall, 80% of the genetic variance is distributed between these clusters and 20% occurs within these clusters.

Discussion

This paper investigates how a population genetics approach can be applied in the context of identification for unknown remains recovered along the United States-México border to the study of variation for persons of Latin American origin from the STRs used for individuation in forensic genetics. This is the first study, to our knowledge, that directly compares the genetic data of deceased immigrants from the two most critical border regions, Arizona and Texas, and investigates their relationship to samples of living ethno-geographic communities within México

and across Central/South America. We hypothesized that the genetic structure analysis of these CODIS STR data would reveal distinctive patterns that provide information on population relationships and geographic origins that are of value to forensic case investigations, identifications, and repatriations. Our results support this supposition, indicating that while PCOME and OpID cases share similarities with the HEA samples (Figure 2), the associations reveal nuanced relationships such that Arizona and Texas deceased immigrants more strongly correspond with Mexican and Central American HEA samples, respectively. The nature of these relationships imply that the Arizona and Texas immigrant groups generally represent two different combinations of source populations. This distinguishability, as well as the patterning of associations, agrees with prior population genetics research (Moreno-Estrada et al. 2014; Rangel-Villalobos et al. 2016; Rubi-Castellanos et al. 2009; Salazar-Flores et al. 2015; Wang et al. 2008), and specifically studies focused on morphogenetic variation among deceased immigrants recovered along the border (Algee-Hewitt 2016; Algee-Hewitt et al. 2018; Hughes et al. 2013; Hughes et al. 2017; New 2018; Spradley 2014; Spradley 2016;). These results also concur with apprehension numbers, as well as expectations for associations based on shared population history (Salzano and Sans 2014; United States Border Patrol 2019b). Lastly, our findings provide additional support for the utility of forensically relevant STRs in ancestry inference and show important promise for future integration into forensic anthropological casework protocols in this Border context (Algee-Hewitt et al. 2016; Hughes et al. 2018). In the section that follows, we provide a detailed accounting of these relationships, the implications of this work for forensic anthropological casework in the Border context, and conclude with our thoughts on the future directions.

Major Trends in Latin American Genetic Variation

Mexican, and Latin American peoples more broadly, are highly morphogenetically diverse and this diversity is patterned in ways that reflect known population histories; studies on the patterning of genetic variation among HEA groups within México identify a North-to-South gradient of admixture with European ancestry being most prominent in the North and Indigenous ancestry being more prominent towards the South (Rubi-Castellanos et al. 2009; Rubi-Castellanos et al. 2009; Salazar-Flores et al. 2015). Indigenous groups within México, however, are genetically differentiated both from each other as well as modern Mexican groups that exhibit less Indigenous admixture, regardless of present-day geographic proximity (Moreno-Estrada et al. 2014; Rangel-Villalobos et al. 2013; Rubi-Castellanos et al. 2009; Salazar-Flores et al. 2015; Wang et al. 2008). This pattern reflects the Spanish colonial history of the region, whereby genetic variation is structured by the effects of European admixture and the sociogeographic isolation of Indigenous peoples (Moreno-Estrada et al. 2014; Rangel-Villalobos et al. 2013; Rubi-Castellanos et al. 2009; Wang et al. 2008). Our results recapitulate these findings in the genetically distinct relationships for the majority of indigenous groups and the strength of Mexican HEA relationships detailed below, validating the observed relationships among populations in our analysis and attesting to the value of the STRs for estimating genetic relationships.

Because literature that assesses the variation in STRs within and between other non-Mexican Central/South American countries is far less frequent, understanding the relationships between OpID or PCOME cases and non-Mexican groups is more challenging. Wang et al. (2008) expand their emphasis beyond México to provide admixture mapping of HEA and Indigenous groups throughout Central and South America, but largely interpret their results in

reference to the breakdown of Indigenous/European components rather than as a direct comparison of the relationship between HEA groups. Salazar-Flores et al. (2015) mostly focuses on the genetic variation of Mexican and Caribbean populations with less Indigenous admixture, but also document genetic similarities between HEA groups from México, Guatemala, El Salvador, and Honduras.

In line with these studies, our analyses demonstrate similarity among El Salvador, Guatemala, and the majority of Mexican HEA and PCOME groups ($R^2 > 82.4\%$). The Honduras sample does not share a strong ($R^2 > 80\%$) relationship with any sample from México, but does demonstrate a strong relationship with Costa Rica ($R^2 = 91\%$). Salazar-Flores and colleagues (2015) also indicate that populations with greater African components to their admixture, such as Black Garifuna from Honduras, Costa Rican, and Caribbean groups, exhibited greater difference from other populations in their analyses. Their results suggest that admixture components better represent the relationship between populations in Central/South America and the Caribbean as variation is not always directly reflective of geographic distance. Our results concur and indicate that geographic proximity alone is not enough to capture the strength and significance of population relationships within Central/South America. Rather, accounting for the distinct, complex patterns of admixture produced by pre-Columbian Indigenous dispersion and the effects of European colonialism can prove invaluable for teasing apart specific genetic signatures for communities in Central American (Moreno-Estrada et al. 2014; Salazar-Flores et al. 2015; Wang et al. 2008).

Arizona-México Immigrant Genetic Variation

The PCOME samples include identified individuals from Northwest México, Central México, Southeast México, and unidentified individuals whose regions of origin are not yet known. Here we explore whether the genetic data of identified individuals correlates more strongly with regionally proximate populations within México. In doing so, we investigate whether it is likely that many individuals remain unidentified because they do not correspond with these regions of México or originate from other regions/countries , whose structural systems may impede family reporting and/or lack support for families of the missing.

In our analysis, the PCOME samples exhibit genetic relationship trends that differ from the OpID samples (Figure 3). While the Northwest PCOME sample shares significant relationships ($p < 0.0001$) with all Mexican HEA groups included in this analysis, no coefficient of determination (R^2) exceeds 90% (Table 2). The Central PCOME sample, however, has relationships greater than 90% for all Central Mexican HEA groups. These results indicate a stronger relationship between the Central PCOME sample and other Central Mexican HEA groups. Notably, all other Central Mexican HEA groups share coefficients greater than 95.4%. Therefore, these results demonstrate good concordance with expected genetic relationships based on geographic proximity.

The PCOME Southeast sample shares strong relationships with only two groups included in this analysis: the PCOME unidentified individuals ($R^2 = 83.5\%$) and the Indigenous Choles sample from South México ($R^2 = 80\%$). This, in conjunction with the sample's independent cluster (Figure 3), indicate a weaker genetic relationship of the PCOME Southeast sample to other Mexican and Central American groups. Hughes et al. (2017) observed similar results that demonstrate individuals from the PCOME Southeast sample have a greater proportion of Indigenous ancestry than members of the other PCOME groups.

The PCOME sample of unidentified individuals has the strongest relationship with identified Central Mexican immigrants (R^2 = 94.1%) and a Mexican HEA sample from the southern state of Guerrero (R^2 = 90.1%). However, as with the Northwest and Central PCOME, the coefficients for all Mexican HEA groups and the Indigenous Choles also remain high (>80%). We have seen thus far in our analysis that, for the majority of groups within México, with closer geographic proximity also comes stronger relationships (>90%). Therefore, it is possible that the stronger relationship between the unidentified PCOME, Central PCOME, and the Guerrero sample may indicate that more of the unidentified individuals share a similar genetic relationship with people from these regions. Alternatively, if the genetic variation captured by the genetic markers runs along a gradient (e.g. Rubi-Castellanos et al. 2013), then it could be that the unidentified PCOME sample represents individuals from all regions of Mexico, that when combined, are presumably most comparable to the Central Mexican group.

Temporal shifts in the demographics of the immigrating populations, and the impact of these changes on genetic structure in the long-term aggregate of cases at PCOME, has been already documented (Algee-Hewitt et al. 2018; Hughes et al. 2017). Hence, it is important to clarify that we do not account for change over time because the data necessary for conducting parallel analyses with OpID is not currently available. We can posit that in failing to partition out the PCOME analysis by temporal cohorts, we are effectively averaging patterns of high European admixture, as is prevalent in older cases, with those indicative of high Indigenous ancestry, prevalent in more recent cases. Thereby, masking potential associations between some of the PCOME temporal cohorts with Central and South American groups. Additionally, PCOME identified individuals from other Central American countries have yet to be

incorporated into our analysis. Further research is needed to address these issues in our comparative Arizona-Texas framework.

Texas-México Immigrant Genetic Variation

Due to small sample size, OpID unidentified and identified individuals were collapsed to create a generalized Texas-México immigrant sample that includes individuals of unknown regions of origin, as well as identified individuals from Ecuador, El Salvador, Guatemala, Honduras, México, and Nicaragua. If the distribution of unidentified individuals reflects a similar distribution to the stated regions of origin for the identified individuals in our sample, then we would expect this sample to share the strongest genetic relationships with El Salvador and México. However, if the distribution of the OpID sample reflects distributions more similar to the 2012 Rio Grande Valley USBP apprehensions (Figure 1), we would expect the OpID sample to share the strongest genetic relationships with México followed by Guatemala, El Salvador, and then Honduras. Our analyses indicate that neither of these scenarios demonstrate the best fit to the OpID genetic data.

OpID shares strong genetic relationships ($R^2 > 80\%$) to groups from Belize, Colombia (Caldas), Costa Rica, El Salvador, Guatemala, Honduras, and Nicaragua. However, only Belize, Honduras, and Nicaragua have coefficients greater than 90%. These results demonstrate considerable difference between OpID and PCOME, as well as the distribution of current identified persons within the OpID sample and 2012 USBP apprehension numbers due to the absence of strong association to any groups from México. Rather, they exhibit greater similarity to the 2019 USBP apprehension numbers. Thus, our results infer that there is a significant underrepresentation of identifications particularly from Nicaragua and Honduras. This suggests

that a greater number of the OpID unidentified individuals may be from geographically

proximate regions to Belize, Honduras, and Nicaragua or communities that similarly share

stronger population relationships. Finally, USBP apprehension distributions suggest that there

may be a temporal component to identified and unidentified OpID cases as well. Whether those

trends are reflected in the genetic structure of OpID cases has yet to be explored and signals

future directions for research supporting the identification of Texas-México immigrants.

Additionally, we emphasize the inverse relationships exhibited by PCOME and OpID

cases. Except for Guatemala and El Salvador, discussed in further detail below, the comparisons

between the PCOME samples and Central/South American populations fall below 67%; whereas

the OpID sample exhibits no genetic associations greater than 71% for all PCOME or Mexican

populations despite having numerous identified individuals from México within the sample.

These results demonstrate that while there is shared genetic variation, the overall genetic

variation between these groups are less predictive of each other. Therefore, we hypothesize that

the majority of individuals within this unidentified PCOME sample are less likely to originate

from Central and South American regions; while the majority of the unidentified individuals

within the OpID sample are less likely to originate from regions within México.

Importance of Indigenous Parental Proxy Sample Selection

It is important to clarify the genetic relationships observed for the HEA groups from Guatemala

and El Salvador, as well as the indigenous Choles group because of their implications for

predicting possible regions of origin in immigrant remains. The El Salvadoran and Guatemalan

samples share strong relationships ($R^2 > 80\%$) with each other, OpID and PCOME immigrants,

Mexican HEA groups, the Choles, and Nicaraguan groups. The Choles only share a strong

relationship ($R^2 > 83.4\%$) with the El Salvador, Guatemala, PCOME immigrants, and Mexican HEA groups. These relationships are important because they demonstrate the limitations of analyses that examine overall genetic variation without breaking down shared genetic admixture components, with particular attention to historically relevant parental populations.

Each of the populations referenced above are inextricably linked by shared pre-Colombian population histories, as well as colonial experiences and histories. For example, Ibarra-Rivera et al. (2008) focused on comparing modern Indigenous groups from México, Guatemala, and El Salvador and found that Mayan groups from México and Guatemala are more genetically similar to each other than they are to other non-Mayan Mesoamerican groups, though there is some shared genetic similarity with the Panchimalco from El Salvador and Kichwas from Ecuador. Additionally, they observed that the Mayan derived Indigenous Choles, exhibited significant admixture with the non-Mayan Salvadoran Indigenous groups. Therefore, the authors argue that although each Indigenous community inhabits geographically "distant and distinct" areas and are often linguistically discrete, the far reaching trade of the Maya and the lack of significant geographic barriers have introduced homogeneity between Maya sourced indigenous groups and non-Mayan indigenous groups that were under Mayan dominion (Ibarra-Rivera et al. 2008). Similarly, Wang et. al (2008) extends this notion into the modern HEA groups by demonstrating correlations between HEA samples from Central México and Guatemala that are linked to pre-Colombian linguistic affiliations.

The relationship between these populations long exceeds the establishment of the modern nation states that we now recognize as México, Guatemala, & El Salvador – thus shared genetic structure between samples from these countries is not unexpected. In fact, the genetic structure observed in these analyses suggest that the Choles may serve as the most informative parental

proxy for Indigenous admixture in future analyses that explore Mexican, Guatemalan, and El Salvadoran genetic variation. Future analyses will verify these hypotheses by investigating more deeply the admixture history and current distribution of proportions for each of these groups, as well as assessing the most informative CODIS STR loci for the purpose of region of origin prediction. The present analysis is significant for the fact that it demonstrates the nuances of these population relationships and the promise for this genetic data to function predictively on the region level.

Implications for Casework and Future Analyses

Who was the unknown in life? How can we find their next of kin? What agency, organization, or institutions should we contact to initiate repatriation? These are the fundamental questions facing forensic practitioners working in context of the U.S.-México border. When thousands of people are attempting to seek refuge by crossing the Southern border and hundreds-to-thousands of those same people are dying each year as a result, reaching families by resolving the issue of pin-pointing home countries carries an incredible amount of weight. Typically, prediction methods for region of origin from genetic data have been reserved for historical/ancient populations, hidden within the black box of large genetics sequencing companies, and/or have required techniques and data not currently available to the average forensic practitioner for reasons of costs, access, and training/expertise. While the analyses presented here still require a specific skillset and would be enhanced by additional data, the predictive potential of forensically significant STRs and the value of mobilizing population genetic analyses as demonstrated here outweighs what we believe are but temporary limitations.

The intention of this study was to introduce a new pathway of investigation for resolving longstanding issues and addressing emerging challenges as the Border crisis evolves with time, under escalating political and social stressors. To this goal, our analyses provide an alternative source of genetic, and specifically "forensic," information for better understanding diversity within Latin America. We have shown not only how work of this nature can corroborate prior research, but also how forensically significant STRs can capture Latin American population variation well, with the potential for accurate prediction on the group or regional level when appropriate source populations are used. This work also expands our knowledge of the patterns of variation in the population of Border fatalities, speaking, for the first time, to the points of genetic similarity and difference between the Arizona and Texas immigrant groups. Lastly, by interpreting our genetic data in light of demographic (e.g., apprehension and migration) information, we are able to convincingly demonstrate how this sample of Arizona-México and Texas-México immigrants represent source populations from two different regions of Central America.

All of these results are meaningful for PCOME and OpID casework because they can help inform their investigative processes: most significantly, this paper's approach can provide valuable information on geographic origins that can, in turn, assist with the critical challenge of locating next of kin. These family searches are the first step towards obtaining DNA reference samples against which unidentified profiles are compared for exclusion or positive identification.

Conclusion and Future Directions

While this work is encouraging, we conclude with some responsible words of caution. Our results speak to general group trends; they cannot yet speak to shared genetic variation on the

individual level. Previous research has attempted to cross-classify individuals from Latin American communities utilizing admixture proportions drawn from forensically significant STRs, but found significant original-to-predicted group deviations (Hughes et al. 2018). There is still considerable exploration to be done to identify the most appropriate methods, the optimal set of genetic markers, and the best parental/reference samples for making predictions on the individual level. It has already been shown that information on ancestry increases as we increase the number of forensic STRs, and that different markers and systems, including other kinds of biological or cultural data, not only support but differently enrich results (Algee-Hewitt et al. 2016; Algee-Hewitt et al. 2018; Algee-Hewitt et al. 2020; Ammer et al. 2020; Bartelink et al. 2020; Hughes et al. 2017; Kramer et al. 2020; Soler et al. 2019). It is likely, then, that our best solutions will utilize a broadened set of genomic (STR, single nucleotide polymorphisms, mitochondrial and Y-STR haplogroups), skeletal, morphological, isotopic, and case context data. We stress that regardless of the specific data or methodologies utilized, it is clear that any region of origin predictions for unidentified groups or individuals will require carefully nuanced, population specific approaches that center Indigenous population histories and the unique lived experience of the individual.

We also call attention to the fact that the present research attempts to address the genetic variation of Latinx immigrants crossing in the Arizona-México and Texas-México regions by using what data are, at present, available. While there are thousands of recovered cases, there are likely many more who will never be found and their unique genetic, morphological, and cultural signatures, therefore, not incorporated into these kinds of analysis. Importantly, there is good reason to believe that some of the unidentified remains may very well not match the "typical" Latin American profile at all (Anderson 2008). In fact, 2019 apprehension numbers indicate

increasing regional diversity, with demographic reports including immigrants from countries such as Bangladesh, Brazil, China, and India (United States Border Patrol 2019b).

Future work needs to focus on both establishing the "typical" genetic profile for the Border dead, as a baseline for comparison, and a flexible framework for defining the immigrant profile as this population continues to diversify. Our future research, therefore, aims to expand our study of Border context deaths to include an admixture-based analysis that delivers individual-level ancestry proportions, as well as addressing the challenges associated with Border deaths representing non-Latinx peoples through the application of unsupervised models that allow for the input of many different data types without making assumptions or requiring prior information about the number of groups in the data or the origins of the individual (Algee-Hewitt 2016; Algee-Hewitt et al. 2018; Algee-Hewitt et al. 2020).

The ability to identify and return immigrants who have lost their lives along the Border to their loved ones is not only dependent upon the kinds of resources available to forensic case practitioners working in the different border regions of Arizona and Texas. It is also dependent on having depth of knowledge in human variation and breadth of academic research on the peoples crossing and dying along the Border to assist in developing new methods for meeting forensic anthropological challenges and responding to changes in practice. Providing opportunities for alternative pathways to reaching families and communities affected by the Border crisis should be a research priority as it has great potential for improving identification rates for these unknowns. While the analyses presented here represent but a small first step toward this much greater goal, it does demonstrate the possibility, the utility, and the necessity for further exploration into the genetic diversity of immigrants who have perished in the Arizona and Texas Border regions. The most successful approach is one that operates under truly multi-

disciplinary framework that draws upon these and other recent advancements in biological

research, and incorporates insights on socio-cultural factors from the anthropologists, human

rights advocates, and aid workers who best understand the people and places affected by the

current immigration crisis and the conditions specific to trans-border contexts (Reineke 2019;

Soler et al. 2019; Spradley and Gocha 2020).

Received 24 October 2020; accepted for publication 18 November 2020.

Literature Cited

Addinsoft. 2020. *XLSTAT*. New York: Addinsoft. https://www.xlstat.com.

Algee-Hewitt, B. F.B. 2016. Population inference from contemporary American craniometrics. *Am. J. Phys. Anthropol.* 160:604–624.

Algee-Hewitt, B. F.B. 2017. Geographic substructure in craniometric estimates of admixture for contemporary American populations. *Am. J. Phys. Anthropol.* 164:260–280.

Algee-Hewitt, B. F.B., C. Coelho, D. Navega et al. 2020. Statistical approaches to ancestry estimation: New and established methods for the quantification of cranial variation for forensic casework. In *Statistics and Probability in Forensic Anthropology*, Z. Obertová, A. Stewart, and C. Cattaneo, eds. London: Academic Press, 227–247.

Algee-Hewitt, B. F.B., M. D. Edge, J. Kim et al. 2016. Individual identifiability predicts population identifiability in forensic microsatellite markers. *Curr. Biol.* 7:935–942.

Algee-Hewitt, B. F.B., C. E. Hughes, and B. Anderson. 2018. Temporal, geographic and identification trends in craniometric estimates of ancestry for persons of Latin American origin. *Forensic Anthropol.* 1:4–17.

Algee-Hewitt, B. F.B., R. Vandewalle, and B. E. Anderson. 2020. The biogeographic profile in forensic anthropology: An alternative to ancestry for casework along the United States–Mexico border. In *American Academy of Forensic Sciences, February 11–22, 2020, Anaheim, CA, USA*.

Ammer, S. T. M., E. J. Bartelink, J. M. Vollner et al. 2020. Spatial distributions of oxygen stable isotope ratios in tap water from Mexico for region of origin predictions of unidentified border crossers. *J. Forensic Sci.* 65:1,049–1,055.

Anderson, B. E. 2008. Identifying the dead: Methods utilized by the Pima County (Arizona) Office of the Medical Examiner for undocumented border crossers: 2001–2006. *J. Forensic Sci.* 53:8–15.

Anderson, B. E., and B. O. Parks. 2008. Symposium on border crossing deaths: Introduction. *J. Forensic Sci.* 53:6–7.

Anderson, B. E., and M. K. Spradley. 2016. The role of the anthropologist in the identification of migrant remains in the American Southwest. *Acad. Forensic Pathol.* 6:432–438.

Barrot, C., C. Sánchez, M. Ortega et al. 2005. Characterisation of three Amerindian populations from Hidalgo State (Mexico) by 15 STR-PCR polymorphisms. *Int. J. Legal Med.* 119:111–113.

Bartelink, E. J., L. A. Chesson, B. J. Tipple et al. 2020. Multi-isotope approaches for region-of-origin predictions of undocumented border crossers from the US-Mexico border: Biocultural perspectives on diet and travel history. In *Forensic Science and Humanitarian Action*, R. C. Parra, S. C. Zapico, and D. H. Ubelaker, eds. Hoboken, NJ: Wiley, 369–384.

Bernal, L. P., L. Borjas, W. Zabala et al. 2006. Genetic variation of 15 STR autosomal loci in the Maracaibo population from Venezuela. *Forensic Sci. Int.* 161:60–63.

Birkby, W. H., T. W. Fenton, and B. E. Anderson. 2008. Identifying Southwest Hispanics using nonmetric traits and the cultural profile. *J. Forensic Sci.* 53:29–33.

Butler, J. M., and C. R. Hill. 2012. Biology and genetics of new autosomal STR loci useful for forensic DNA analysis. *Forensic Sci. Rev.* 24:15–26.

Cornelius, W. A. 2001. Death at the border: Efficacy and unintended consequences of US immigration control policy. *Popul. Dev. Rev.* 27:661–685.

Dudzik, B., and R. L. Jantz. 2016. Misclassifications of Hispanics using Fordisc 3.1: Comparing cranial morphology in Asian and Hispanic populations. *J. Forensic Sci.* 61:1,311–1,318.

Flores, S., J. Sun, J. King et al. 2015. Allele frequencies for 15 autosomal STR loci and haplotype data for 17 Y-STR loci in a population from Belize. *Int. J. Legal Med.* 129:1,217–1,218.

Gocha, T., M. K. Spradley, and R. Strand. 2018. Bodies in limbo: Issues in identification and repatriation of migrant remains in South Texas. In *Sociopolitics of Migrant Death and Repatriation: Perspectives from Forensic Science*, K. E. Latham and A. J. O'Daniel, eds. Cham, CH: Springer International Publishing, 143–156.

González-Martín, A., A. Gorostiza, H. Rangel-Villalobos et al. 2008. Analyzing the genetic structure of the Tepehua in relation to other neighbouring Mesoamerican populations. A study based on allele frequencies of STR markers. *Am. J. Hum. Biol.* 20:605–613.

Gorostiza, A., A. González-Martín, C. López Ramírez et al. 2007. Allele frequencies of the 15 AmpF/Str Identifiler loci in the population of Metztitlán (Estado de Hidalgo), México. *Forensic Sci. Int.* 166:230–232.

Gutiérrez, G. 2015. Identity erasure and demographic impacts of the Spanish caste system on the Indigenous populations of México. In *Beyond Germs: Native Depopulation in North America*, C. M. Cameron, P. Kelton, and A. C. Swedlund, eds. Tucson, AZ: University of Arizona Press, 119–145.

Haasl, R. J., and B. A. Payseur. 2011. Multi-locus inference of population structure: A comparison between single nucleotide polymorphisms and microsatellites. *Heredity (Edinb.)* 106:158–171.

Hefner, J. T., M. A. Pilloud, C. J. Black et al. 2015. Morphoscopic trait expression in "Hispanic" populations. *J. Forensic Sci.* 60:1,135–1,139.

Hughes, C. E., B. F.B. Algee-Hewitt, and L. Konigsberg. 2018. Population identifiability from forensic genetic markers: Ancestry variation in Latin America. *Hum. Biol.* 90:161–175.

Hughes, C. E., B. F.B. Algee-Hewitt, R. Reineke et al. 2017. Temporal patterns of Mexican migrant genetic ancestry: Implications for identification. *Am. Anthropol.* 119:193–208.

Hughes, C. E., B. Dudzik, B. F.B. Algee-Hewitt et al. 2019. Understanding (mis)classification trends of Latin Americans in Fordisc 3.1: Incorporating cranial morphology, microgeographic origin, and admixture proportions for interpretation. *J. Forensic Sci.* 64:353–366.

Hughes, C. E., M. L. Tise, L. H. Trammell et al. 2013. Cranial morphological variation among contemporary Mexicans: Regional trends, ancestral affinities, and genetic comparisons. *Am. J. Phys. Anthropol.* 151:506–517.

Humane Borders. 2021. www.humaneborders.org.

Ibarra-Rivera, L., S. Mirabal, M. M. Regueiro et al. 2008. Delineating genetic relationships among the Maya. *Am. J. Phys. Anthropol.* 135:329–347.

Juarez, C., R. Ramey, D. T. Flaherty et al. 2020. Utility of stable isotope ratios of tap water and human hair in determining region of origin in Central and Southern Mexico: Modeling relationships between δ^2H and δ^{18}O isotope inputs in modern Mexican hair. In *Forensic Science and Humanitarian Action*, R. C. Parra, S. C. Zapico, and D. H. Ubelaker, eds. Hoboken, NJ: Wiley, 345–367.

Juárez-Cedillo, T., J. Zuñiga, V. Acuña-Alonzo et al. 2008. Genetic admixture and diversity estimations in the Mexican Mestizo population from Mexico City using 15 STR polymorphic markers. *Forensic Sci. Int. Genet.* 2:e37–e39.

Kramer, R. T., E. J. Bartelink, N. P. Herrmann et al. 2020. Application of stable isotopes and geostatistics to infer region of geographical origin for deceased undocumented Latin American migrants. In *Forensic Science and Humanitarian Action*, R. C. Parra, S. C. Zapico, and D. H. Ubelaker, eds. Hoboken, NJ: Wiley, 425–440.

Liu, N., L. Chen, S. Wang et al. 2005. Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genet.* 6:1–5.

Locia-Aguilar, G. J., B. López-Saucedo, S. Deheza-Bautista et al. 2018. Forensic parameters for 15 autosomal STRs in Mestizo population from the state of Guerrero (South, Mexico). *Leg. Med. (Tokyo)* 32:104–106.

Luna-Vazquez, A., G. Vilchis-Dorantes, M. O. Aguilar-Ruiz et al. 2005. Population data for 15 loci (Identifiler® Kit) in a sample from the Valley of Mexico. *Leg. Med. (Tokyo)* 7:331–333.

Ma, L., Y. Ji, and D. Zhang. 2015. Statistical measures of genetic differentiation of populations: Rationales, history and current states. *Curr. Zool.* 61:886–897.

Maier, C. A., and R. L. George. 2020. Examining differences in presumed migrants from Texas and Arizona using cranial and dental data. *Forensic Anthropol.* 3:17–29.

Martinez, D., R. Reineke, R. Rubio-Goldsmith et al. 2013. *A Continued Humanitarian Crisis at the Border: Undocumented Border Crosser Deaths Recorded by the Pima County Office of the Medical Examiner, 1990–2012.* Tuscan, AZ: Binational Migration Institute, University of Arizona.

Martinez-Espin, E., L. J. Martínez-Gonzalez, F. Fernandez-Rosado et al. 2006. Guatemala Mestizo population data on 15 STR loci (Identifiler® Kit). *J. Forensic Sci.* 51:1,216–1,218.

Matamoros, M., Y. Pinto, F. J. Inda et al. 2008. Population genetic data for 15 STR loci (Identifiler™ kit) in Honduras. *Leg. Med. (Tokyo)* 10:281–283.

Meirmans, P. G., and P. W. Hedrick. 2011. Assessing population structure: FST and related measures. *Mol. Ecol. Resour.* 11:5–18.

Morales, J. A., J. C. Monterrosa, and J. Puente. 2004. Population genetic data from El Salvador (Central America) using AmpFlSTR® Identifiler® PCR Amplification Kit. *Int. Congr. Ser.* 1261:223–225.

Moreno-Estrada, A., C. R. Gignoux, J. C. Fernandez-Lopez et al. 2014. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science* 344:1,280–1,285.

New, B. T. 2018. Using matched craniometric and genetic data to assess the population structure of Texas-Mexico border migrants. Master's thesis, Texas State University.

New, B. T., M. K. Spradley, and L. Fehren-Schmitz. 2019. A forensic anthropological perspective on Latin American genetic diversity. In *Program of the 88th Annual Meeting of the American Association of Physical Anthropologists, March 2019, Cleveland, OH, USA*. 175.

Nuñez, C., M. Baeta, C. Sosa et al. 2010. Reconstructing the population history of Nicaragua by means of mtDNA, Y-chromosome STRs, and autosomal STR markers. *Am. J. Phys. Anthropol.* 143:591–600.

Peakall, R., and P. E. Smouse. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic
software for teaching and research—an update. *Bioinformatics* 28:2,537–2,539.

Pima County Office of the Medical Examiner. 2019. *Pima County Office of the Medical
Examiner – Annual Report 2019.* Tuscan, AZ: PCOME.

Porras, L., L. Beltrán, T. Ortiz et al. 2008. Genetic polymorphism of 15 STR loci in central
western Colombia. *Forensic Sci. Int. Genet. 2*:e7–e8.

Rangel-Villalobos, H., V. M. Martínez-Sevilla, G. Martínez-Cortés et al. 2016. Importance of the
geographic barriers to promote gene drift and avoid pre- and post-Columbian gene flow
in Mexican Native groups: Evidence from forensic STR loci. *Am. J. Phys. Anthropol.*
160:298–316.

Rangel-Villalobos, H., V. M. Martínez-Sevilla, J. Salazar-Flores et al. 2013. Forensic parameters
for 15 STRs in eight Amerindian populations from the north and west of Mexico.
Forensic Sci. Int. Genet. 7:e62–e65.

Reineke, R. C. 2019. Necroviolence and postmortem care along the US-México border. In *The
Border and Its Bodies: The Embodiment of Risk Along the U.S.-México Line*, T. E.
Sheridan and R. H. McGuire, eds. Tuscan, AZ: University of Arizona Press, 144–282.

Reineke, R. C., and C. Halstead. 2017. Identifying dead migrants: Examples from the United
States–Mexico border. In *Fatal Journeys, Volume 3, Part 1: Improving Data on Missing
Migrants*, F. Laczko, A. Singleton, and J. Black, eds. Geneva, CH: International
Organization for Migration, 77–98.

Rodríguez, A., G. Arrieta, I. Sanóu et al. 2007. Population genetic data for 18 STR loci in Costa
Rica. *Forensic Sci. Int.* 168:85–88.

Rosenberg, N. A., L. M. Li, R. Ward et al. 2003. Informativeness of genetic markers for inference of ancestry. *Am. J. Hum. Genet.* 73:1,402–1,422.

Rubi-Castellanos, R., M. Anaya-Palafox, E. Mena-Rojas et al. 2009. Genetic data of 15 autosomal STRs (Identifiler kit) of three Mexican Mestizo population samples from the states of Jalisco (West), Puebla (Center), and Yucatan (Southeast). *Forensic Sci. Int. Genet.* 3:e71–e76.

Rubi-Castellanos, R., G. Martínez-Cortés, J. F. Muñoz-Valle et al. 2009. Pre-Hispanic Mesoamerican demography approximates the present-day ancestry of Mestizos throughout the territory of Mexico. *Am. J. Phys. Anthropol.* 139:284–294.

Rubio-Goldsmith, R., M. McCormick, D. Martinez et al. 2006. *The 'Funnel Effect' & Recovered Bodies of Unauthorized Migrants Processed by the Pima County Office of the Medical Examiner, 1990–2005*. Tuscan, AZ: Binational Migration Institute, University of Arizona.

Salazar-Flores, J., F. Zuniga-Chiquette, R. Rubi-Castellanos et al. 2015. Admixture and genetic relationships of Mexican Mestizos regarding Latin American and Caribbean populations based on 13 CODIS-STRs. *Homo.* 66:44–59.

Salzano, F. M., and M. Sans. 2014. Interethnic admixture and the evolution of Latin American populations. *Genet. Mol. Biol.* 37:151–170.

Sánchez, C., C. Barrot, M. Ortega et al. 2005. Genetic diversity of 15 STRs in Choles from northeast of Chiapas (Mexico). *J. Forensic Sci.* 50:1–3.

Sánchez-Serrano, C. 1996. Mestizaje e historia de la población en México. In *Polimorfismo Génico (HLA) en Poblaciones Hispanoamericanas*, A. Martín-Municio and P. García-

Barreno, eds. Madrid, ES: Real Academia de Ciencias Exactas, Físicas y Naturales, 173–193.

Soler, A., R. C. Reineke, J. S. Beatrice et al. 2019. Etched in bone: Embodied suffering in the remains of undocumented migrants. In *The Border and Its Bodies: The Embodiment of Risk Along the U.S.-México Line*, T. E. Sheridan and R. H. McGuire, eds. Tuscan, AZ: University of Arizona Press, 175–207.

Soto, G., and D. E. Martínez. 2018. The geography of migrant death: Implications for policy and forensic science. In *Sociopolitics of Migrant Death and Repatriation*, K. E. Latham and A. J. O'Daniel, eds. Cham, CH: Springer International Publishing, 67–82.

Spradley, M. K. 2014. Toward estimating geographic origin of migrant remains along the United States–Mexico border. *Ann. Anthropol. Pract.* 38:101–110.

Spradley, M. K. 2016. Biological distance, migrants, and reference group selection in forensic anthropology. In *Biological Distance Analysis: Forensic and Bioarchaeological Perspectives*, M. A. Pilloud and J. T. Hefner, eds. London: Academic Press, 231–244.

Spradley, M. K., and T. P. Gocha. 2020. Migrant deaths along the Texas/Mexico border: A collaborative approach to forensic identification of human remains. In *Forensic Science and Humanitarian Action*, R. C. Parra, S. C. Zapico, and D. H. Ubelaker, eds. Hoboken, NJ: Wiley, 535–548.

Spradley, M. K., M. E. Jasinski, J. Arciszewska et al. 2019. Giving them back their names and faces: DNA analyses and material culture studies of undocumented migrants from Central America and Mexico to Texas. In *Program of the Conference of International Society for Forensic Genetics, September 2019, Prague, Czech Republic.* Mainz, DE: International Society for Forensic Genetics, 62.

Tise, M. L., E. H. Kimmerle, and M. K. Spradley. 2014. Craniometric variation of diverse

populations in Florida: Identification challenges within a border state. *Ann. Anthropol.*

Pract. 38:111–123.

United States Border Patrol. 2020a. *U.S. Border Patrol Fiscal Year Southwest Border Sector*

Deaths (FY 1998 - FY 2019). Washington, D.C.: U.S. Customs and Border Protection.

United States Border Patrol. 2020b. *U.S. Border Patrol Nationwide Apprehensions by*

Citizenship and Sector (FY 2007 - FY 2019). Washington, D.C.: U.S. Customs and

Border Protection.

U.S. Customs and Border Protection. 2020. Stats and Summaries.

https://www.cbp.gov/newsroom/media-resources/stats (accessed 2020).

Wang, S., N. Ray, W. Rojas et al. 2008. Geographic patterns of genome admixture in Latin

American Mestizos. *PLoS Genet.* 4:e1000037.

Table 1. Population Data with Population Names, Codes Used in Analyses, Sample Sizes, Admixture Proportion Designations, Geographic Regions of Origin, and References

Population Name	Code	n	Indigenous or HEA	Region	Reference
Texas-México immigrants	OpID	61	Unknown	Unknown	Present Study
Arizona-México immigrants w/ ID	PIDNW	32	Unknown	Northwest México	Hughes *et al.* 2017
Arizona-México immigrants w/ ID	PIDCen	45	Unknown	Central México	Hughes *et al.* 2017
Arizona-México immigrants w/ ID	PIDSE	27	Unknown	Southeast México	Hughes *et al.* 2017
Arizona-México immigrants w/o ID	PUBC	238	Unknown	Unknown	Hughes *et al.* 2017
Belize	Bel	290	HEA	Belize	Flores *et al.* 2015
Caldas	Cald	1212	HEA	Colombia	Porras *et al.* 2008
Quindio	Quin	80	HEA	Colombia	Porras *et al.* 2008
Risaralda	Risa	652	HEA	Colombia	Porras *et al.* 2008
Costa Rica	CosRi	496	Probable HEA	Costa Rica	Rodríguez *et al* 2007
El Salvador	ElSalv	228	Probable HEA	El Salvador	Morales *et al.* 2004
Guatemala	Guat	200	HEA	Guatemala	Martinez-Espin *et al.* 2006
Honduras	Hond	198	HEA	Honduras	Matamoros *et al.* 2008
Otomi SM	OtoSM	91	Indigenous	East México	Barrot *et al.* 2005
Otomi XM	OtoXM	83	Indigenous	East México	Barrot *et al.* 2005
Huasteco	Huas	135	Indigenous	East México	Barrot *et al.* 2005
Tepehua	Tph	57	MA	East México	González-Martín *et al.* 2008
Metztitlán	Metz	180	HEA	East México	Gorostiza *et al.* 2007
México City	MexCit	378	HEA	Central México	Juárez-Cedillo *et al.* 2008

Valley of México	ValMex	242	HEA	Central México	Luna-Vazquez *et al.* 2005
Tarahumara	Tar	204	Indigenous	North México	Rangel-Villalobos *et al.* 2013
Mayos	Myo	45	Indigenous	North México	Rangel-Villalobos *et al.* 2013
Huichol	Hui	239	Indigenous	West México	Rangel-Villalobos *et al.* 2013
Tepehuano	Tep	123	Indigenous	West México	Rangel-Villalobos *et al.* 2013
Cora	Cora	85	Indigenous	West México	Rangel-Villalobos *et al.* 2013
Mexicaneros	Mex	84	Indigenous	West México	Rangel-Villalobos *et al.* 2013
Choles	Chol	109	Indigenous	South México	Sánchez *et al.* 2005
Guerrero	Guer	251	HEA	South México	Locia-Aguilar *et al.* 2018
Nicaragua	Nica	163	HEA	Nicaragua	Nuñez *et al.* 2010
Venezuela	Venez	203	Probable HEA	Venezuela	Bernal *et al.* 2006

Table 2. Spearman Correlation Coefficients of Determination Matrix

Only groups with statistically significant relationships (p < 0.0001) and coefficients of determination (R^2) greater than 0.80 to either OpID or PCOME are depicted. $R^2 > 0.80$ are highlighted in bold and green.

Variables	OpID	PUBC	PIDNW	PIDCen	PIDSE	Belize	Caldas	Costa Rica	El Salvador	Guatemala	Honduras	Metztitlan	Mex City	Guerrero	Valley Mex	Choles	Nicaragua
OpID	1	0.683	0.637	0.671	0.529	**0.913**	**0.814**	**0.828**	**0.864**	**0.8**	**0.913**	0.618	0.666	0.709	0.685	0.659	**0.947**
PUBC	0.683	1	**0.89**	**0.941**	**0.835**	0.537	0.46	0.467	**0.851**	**0.824**	0.572	**0.859**	**0.882**	**0.901**	**0.895**	**0.876**	0.669
PIDNW	0.637	**0.89**	1	**0.844**	0.745	0.566	0.439	0.481	**0.817**	0.777	0.585	**0.862**	**0.886**	**0.864**	**0.866**	**0.834**	0.656
PIDCen	0.671	**0.941**	**0.844**	1	0.749	0.514	0.445	0.435	**0.88**	**0.864**	0.576	**0.928**	**0.942**	**0.955**	**0.954**	**0.915**	0.649
PIDSE	0.529	**0.835**	0.745	0.749	1	0.412	0.334	0.335	0.689	0.651	0.419	0.691	0.715	0.753	0.715	**0.799**	0.501
Belize	**0.913**	0.537	0.566	0.514	0.412	1	**0.905**	**0.925**	0.71	0.636	**0.926**	0.473	0.519	0.539	0.523	0.496	**0.901**
Caldas	**0.814**	0.46	0.439	0.445	0.334	**0.905**	1	**0.946**	0.595	0.571	**0.864**	0.367	0.399	0.422	0.403	0.374	**0.838**
Costa Rica	**0.828**	0.467	0.481	0.435	0.335	**0.925**	**0.946**	1	0.606	0.567	**0.91**	0.386	0.412	0.432	0.416	0.385	**0.873**
El Salv	**0.864**	**0.851**	**0.817**	**0.88**	0.689	0.71	0.595	0.606	1	**0.966**	0.788	**0.838**	**0.885**	**0.916**	**0.899**	**0.891**	**0.857**
Guat	**0.8**	**0.824**	0.777	**0.864**	0.651	0.636	0.571	0.567	**0.966**	1	0.754	**0.838**	**0.869**	**0.896**	**0.885**	**0.882**	**0.827**
Honduras	**0.913**	0.572	0.585	0.576	0.419	**0.926**	**0.864**	**0.91**	0.788	0.754	1	0.549	0.585	0.603	0.59	0.567	**0.969**
Metztitlan	0.618	**0.859**	**0.862**	**0.928**	0.691	0.473	0.367	0.386	**0.838**	**0.838**	0.549	1	**0.975**	**0.954**	**0.967**	**0.938**	0.602
MexCity	0.666	**0.882**	**0.886**	**0.942**	0.715	0.519	0.399	0.412	**0.885**	**0.869**	0.585	**0.975**	1	**0.98**	**0.99**	**0.946**	0.648
Guerrero	0.709	**0.901**	**0.864**	**0.955**	0.753	0.539	0.422	0.432	**0.916**	**0.896**	0.603	**0.954**	**0.98**	1	**0.99**	**0.968**	0.678
ValleyMex	0.685	**0.895**	**0.866**	**0.954**	0.715	0.523	0.403	0.416	**0.899**	**0.885**	0.59	**0.967**	**0.99**	**0.99**	1	**0.956**	0.658
Choles	0.659	**0.876**	**0.834**	**0.915**	**0.799**	0.496	0.374	0.385	**0.891**	**0.882**	0.567	**0.938**	**0.946**	**0.968**	**0.956**	1	0.642
Nicaragua	**0.947**	0.669	0.656	0.649	0.501	**0.901**	**0.838**	**0.873**	**0.857**	**0.827**	**0.969**	0.602	0.648	0.678	0.658	0.642	1

Figure Captions

Figure 1. Counts of apprehensions by citizenship for the Tucson and Rio Grande Valley sectors in FY 2012 and FY 2019 as reported by the USBP (United States Border Patrol, 2019b).

Figure 2. MDS plot depicting intrapopulation relationships. Kruskal's stress = 0.133.

Figure 3. Dendrogram depicting dissimilarity classes. Number of classes = 6.
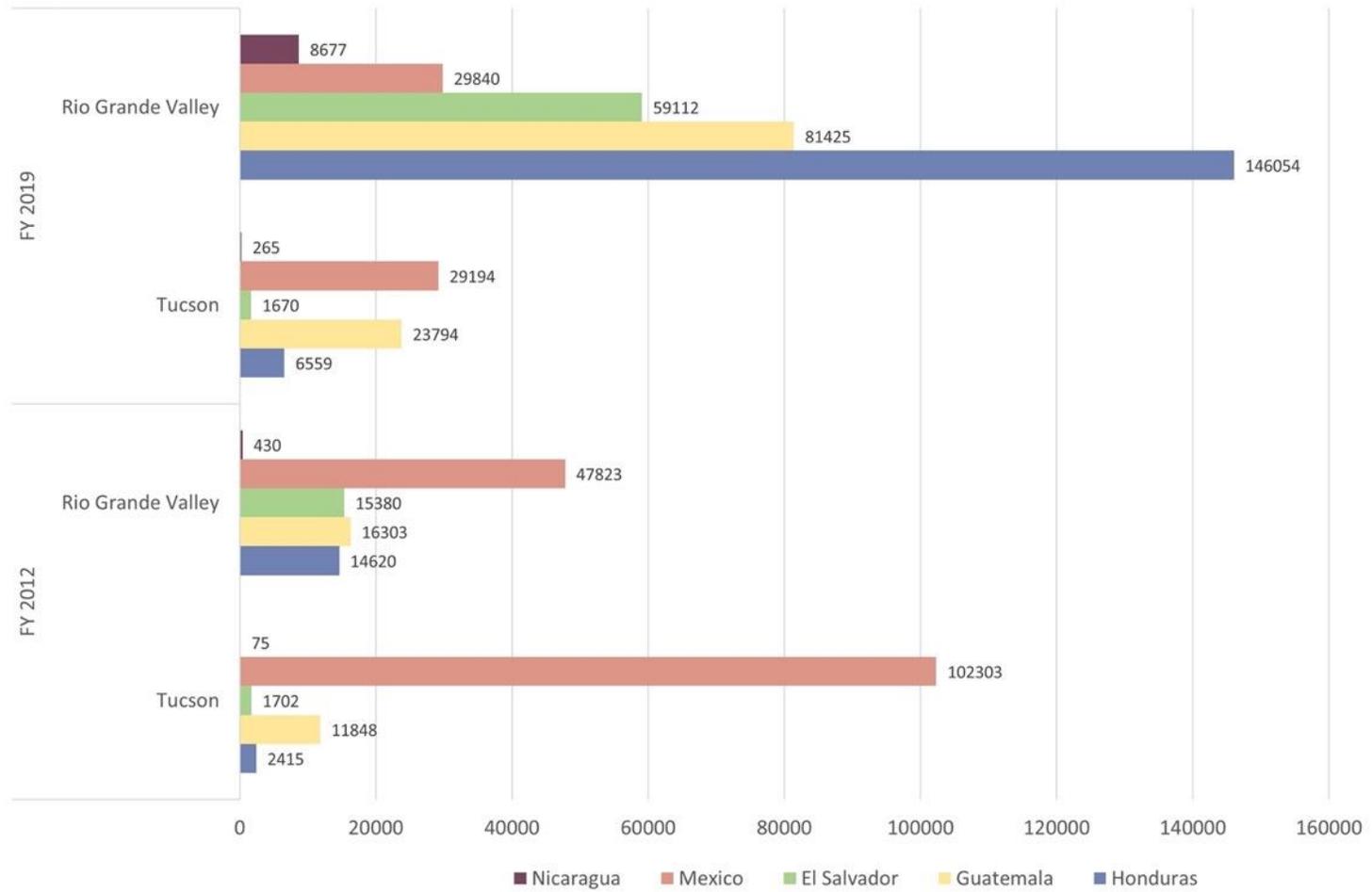
Figure 1.

Figure 2.

Figure 3.