Population affinity estimation on a Spanish sample

Sieber, Katie S; Gómez García-Donas, Julieta

Published in:
Legal Medicine

DOI:
10.1016/j.legalmed.2022.102180

Publication date:
2022

Journal Pre-proofs

Population affinity estimation on a Spanish sample: testing the validity and accuracy of cranium and mandible online software methods

Katie S Sieber, Julieta Gómez García-Donas

PII: S1344-6223(22)00168-7
DOI: https://doi.org/10.1016/j.legalmed.2022.102180
Reference: LEGMED 102180

To appear in: *Legal Medicine*

Received Date: 28 August 2022
Revised Date: 6 November 2022
Accepted Date: 17 November 2022

Please cite this article as: Sieber, K.S., Gómez García-Donas, J., Population affinity estimation on a Spanish sample: testing the validity and accuracy of cranium and mandible online software methods, *Legal Medicine* (2022), doi: https://doi.org/10.1016/j.legalmed.2022.102180

Population affinity estimation on a Spanish sample: testing the validity and accuracy of cranium and mandible online software methods

Katie S Sieber[1], Julieta Gómez García-Donas[1]

[1] Centre for Anatomy and Human Identification, School of Science and Engineering, University of Dundee, Dundee, U.K.

Corresponding author:

Julieta Gómez García-Donas

Email: jgomezgarciadonas001@dundee.ac.uk

MSI Building
Office G L1-541
Dow Street

University of Dundee
Centre for Anatomy and Human Identification
School of Science and Engineering, University of Dundee
Dundee, U.K.

Highlights

- Population affinity methods require validation studies to ensure accurate results
- Online software techniques are a popular choice for population affinity estimation
- AncesTrees and (hu)MANid were used on a contemporary Spanish sample
- AncesTrees using the skull performed better than (hu)MANid using the mandible
- Model selection, reference population and statistical values need to be considered

Abstract

Population affinity estimation is an important step in the identification of unknown individuals. To ensure accurate results, validation studies of newly developed methods must be performed using different target populations and skeletal elements. This research aims to determine the accuracy and reliability of population affinity estimation on a modern Spanish sample using two online software applications.

The sample consists of 114 adult individuals (51 males, 63 females) using 38 measurements and one angle from the skull and mandible. AncesTrees was used for craniometric measurements and (hu)MANid for mandibular variables with different classification models and probability thresholds being evaluated. The required parameters were inputted for each individual and statistics were generated to assess the accuracy of the estimation.

AncesTrees performed with the greatest accuracy as the program correctly classified the sample as Southwestern European or European, with highest accuracies being 54.56% (trial 1), 86.05% (trial 2), 82.61% (trial 3), 34.55% (trial 4) and 100.00% (trial 5). (hu)MANid correctly classified the sample as being from white origin with accuracies ranging from 70.59 to 80.00% without considering correct sex estimation, while accuracy ranged between 62.75 and 80.00% accounting for estimated sex.

Population affinity estimation may determine subsequent methods used in the construction of the biological profile. Our results demonstrate varying accuracy rates depending

on the element and method, offering a critical view in relation to software applicability and validity. Reference populations and intrinsic and extrinsic factors can potentially influence the method accuracy and reliability. Future research should focus on the inclusion of underrepresented groups.

Key words: population affinity, skull, mandible, online software, Spanish sample

1. Introduction

When unknown human skeletal remains are encountered, one of the first steps taken is to establish their identity through the construction of the biological profile [1]. Population affinity, sex, age and stature are the main pieces of information to be gathered for the purpose of forensic identification [2].

The estimation of population affinity has brought some controversy due to its implication to social constructs, with a debate being held on the terminology, sample representation, classification groups, etc. [3]. Several authors have proposed to omit the estimation of population affinity in forensic anthropology reports due to the misconceptions and possible implications for the identification process [4], although others consider this piece of information determinant for positive identification [5]. In practical terms, population affinity might be crucial for the overall analysis performed to build the individual's identity, as some methods for sex and age estimation are population-specific [6-8]. Thus, it is important to accurately estimate population affinity, and to further understand population affinity based on population structure – as proposed by Ross and Pilloud [3] – as this will allow the practitioner to choose the appropriate methods for subsequent assessment [9, 10].

Both morphoscopic and metric approaches have been developed to estimate population affinity [9, 11-16]. Regarding the metric approach, its objectivity based on the nature of data collection has been acknowledged and extensive research has been conducted [11-13, 17, 18]. Moreover, from all the skeletal elements used for population affinity estimation, the cranial skeleton is considered the most suitable as it might not be as environmentally affected as the postcranial elements [10, 19-21]. Thus, the number of metric methods developed using the cranium has increased in the last decade, with techniques using traditional methods employing discriminant function analysis as well as computerized methods being developed from large reference samples [12, 17, 18, 22]. Considering the effect that population affinity has on the ensuing estimations and the emerging number of population affinity estimation methods, validation studies are required to ensure accurate and reliable results [21-25].

The present study aims to explore two computerized population affinity estimation methods developed for the skull and mandible, AncesTrees [17] and (hu)MANid [18], to

determine their reliability and validity to predict group membership of a contemporary Spanish sample [26]. Different prediction models and classification strategies were considered, accounting for different scenarios, and their respective outcomes were evaluated based on data statistics outputs produced by the programs. The results are discussed not only in relation to accuracy rates considering the use of the cranium or mandible for population affinity estimation, but also in relation to the different methodological approaches adopted by each technique.

2. Materials and Methods

2.1. Materials

The materials used for this study consist of the original sample used by Del Río Muñoz [26]. The individuals included in the sample are part of the collection housed at the School of Legal Medicine (SLM) (Universidad Complutense de Madrid, Spain). The SLM collection was acquired through exhumations of cemeteries surrounding Madrid city. The sample is comprised of 114 adult individuals of known sex (51 males, 63 females) who died between 1975 and 1985, and for which age is known for 100 specimens. The mean age for the total sample is 72.7 years, while the mean age at death for males and females is 70.4 years and 74.8 years, respectively. The individuals in the sample are representative of the Spanish population, with individuals from different geographical areas within the country being included [26].

2.2. Methods

The cranial and mandibular data employed in this study was obtained from Del Río Muñoz [26]. The data consists of 38 measurements and one angle, which are required for the application of the methods tested in this study (*see supplementary material*). The cranial and mandibular measurements follow the descriptions from Howells [27] and Martin and Saller [28]. These descriptions are regarded as the gold standard and continue to be used routinely by both experienced and inexperienced scholars [29, 30]. The two-population affinity estimation software that were tested using the modern Spanish sample data were AncesTrees [17] and (hu)MANid [18], which are both freely available online.

AncesTrees was selected due to the software being easily accessible online and its inclusion of a European reference sample [17] (https://osteomics.com/AncesTrees/). The AncesTrees reference dataset includes 1734 individuals (907 males, 827 females) from Howells' Craniometric Series [27, 31, 32]. The biogeographical groups in which an unknown individual can be allocated to when using this software are Northern Asia & Arctic, North America, South America, Southwestern Europe, Northern & Central Europe, Northeast Africa, Sub-Saharan Africa, South Asia, East Asia, Southeast Asia, Polynesia and Australia & Melanesia [17]. The software allows the user to input up to a maximum of 30 measurements. Thus, when the data permitted, the required craniometric values for the respective parameters were manually entered into AncesTrees for each of the 114 individuals in the sample [26]. For consistency and to facilitate better comparison to validation studies already published on this method [21], the left side value was used. When the left-side measurement was not available, the right-side was used instead.

The accuracy rates associated with AncesTrees were evaluated in five trials to test different scenarios, models, and potential biogeographical outcomes. For the first trial, the automated default "tournamentForest" algorithm with the default number of trees (256) was selected because it provides the optimal binary classification and performs best when prior knowledge regarding the unknown individual's population affinity is limited, as might be in a real-life scenario [21, 33]. Following the instructions from the authors [17], the "clustered, 12 clusters" option was selected in order to replicate a scenario in which the population affinity of the individual is unknown. Trial 2 employed the same parameters as trial 1, but with the "clustered, 9 clusters" option selected to determine if the correct classification rates increase when the ancestral classification groups were less specific (e.g., Europe combined as a single reference group instead of Northern/Central and Southern Europe sub-groups). Trials 3 through 5 were conducted following the model parameters employed by Fernandes *et al.* [21] to facilitate comparison to the accuracy rates obtained using AncesTrees on the Brazilian sample. For trial 4 and 5, the "ancestralForest" algorithm was selected to test the software accuracy when allocating the unknown skull to the ancestral clusters selected based on its group membership likelihood [17] (**Table 1**). In trial 4, the "clustered, 9 clusters" option was selected, whereas the

"clustered, 12 clusters" option was chosen for trial 5 with specific selection for Northern & Central Europe, Southwestern Europe, Northeast Africa, and Sub-Saharan Africa groups. The number of trees for these trials was 512 and the number of sub-forests was 32. The boot-strap fraction was set as 63.2% for both trials. Additionally, the trials were processed with balanced bootstrap without replacement and the pseudo-random number generator seed was set at 1989. The "yes" option was selected for parallel computing. For all scenarios, a correct classification was considered when an estimation of Southwestern European or European was produced in the software output, depending on the trial group selection. Correct group classification was first considered regardless of group membership value as in Fernandes *et al.* [21]. Secondly, greater than or equal to 0.80 group membership values were further considered, which indicates the probability of an individual belonging to the ancestral groups selected by the model [17].

Table 1. AncesTrees model parameters for trials 4 and 5.

Trial #	**4**	**5**
Algorithm	**ancestralForest**	**ancestralForest**
Biogeographic Ancestry Coding	*Clustered, 9 clusters*	*Clustered, 12 clusters*
Number of Trees	512	512
Number of Sub-forests	32	32
Bootstrap Fraction	63.2%	63.2%
Balanced Bootstrap	Yes	Yes
Bootstrap with Replacement	No	No
Pseudo-random Number Generator Seed	1989	1989
Parallel Computing	Yes	Yes

* All available ancestries selected.
** Northern & Central Europe, Southwestern Europe, Northeast Africa, Sub-Saharan Africa selected.

(hu)MANid was tested in the present research due to its easy accessibility online and its inclusion of a white (European) reference sample [18] (https://anthropologyapps.shinyapps.io/humanid/). The (hu)MANid reference dataset includes 1745 individuals (1140 males, 505 females) [34]. Based on the nature of the target sample and as recommended by the authors [18], only modern samples from the reference population groups as well as sex outcomes were selected, with the groups being: white male 20th century (WM (20c)), white female 20th century (WF (20c)), black male 20th century (BM (20c)), black female 20th century (BF (20c)), Chinese male (CHM), Hispanic male (HM), Guatemalan male (GUATM), Guatemalan female (GUATF), Cambodian male (CAMM), Cambodian female (CAMF), Vietnamese male (VIETM), Thai male (THAIM), Thai female (THAIF), Korean male (KORM) and Korean female (KORF). When the mandibular data permitted [26], a maximum of nine measurements were entered manually into (hu)MANid for each of the 114 individuals in the sample. (hu)MANid does not specify whether the left- or right-side measurement should be used for bilateral parameters however, the left-side measurement was chosen for consistency. If the left side measurement was not available, then the right side was used. The "more than two groups" option was chosen to compare the sample to the groups included in the dataset to replicate a real-life scenario in which limited background information from the individual is known. Multiple discriminant analysis (MDA) was selected as it has been demonstrated to produce, on average, accuracy rates 9.30% greater than linear discriminant function analysis (LDFA) [18]. Forward Wilks was used in this study, as according to the authors it produced an optimal performance when compared to other stepwise procedures [18]. Moreover, posterior probability (PP) was considered as a measure of the likelihood of an unknown individual belonging to a particular group [33]. To facilitate comparison to FORDISC3 [13] and to conform with the information provided by Elliott and Collard [23], PP values greater than 0.5 and 0.8 were explored, separately, considering that a result below these thresholds is likely to be an incorrect classification [23]. In our study, in the first instance correct classification was considered to be WM (20c) or WF (20c) regardless of sex. Further analysis included correct classification for both population affinity and sex.

Data was analyzed by gathering the information about group membership obtained through the different classification strategies for each software. The number of correctly classified individuals was converted into group classification accuracy percentages. For example, the ancestral group classification was considered to be correct when the unknown individual was classified as European or White. Moreover, the incorrect ancestral group membership percentages were also calculated to further explore the different geographical groups in which the Spanish individuals were allocated to. This was recorded to understand any possible relationship between the reference and target samples. Finally, when the methods allowed for sex-specific population affinity estimation, as in (hu)MANid, both correct population affinity and sex estimation were also recorded.

3. Results

The descriptive statistics for the cranial and mandibular measurements and the mandibular angle employed in this research [26] can be found in *supplementary material*.

3.1. AncesTrees

Out of the 114 individuals in the sample, a total of 110 individuals had the required measurements to use AncesTrees. The classification rates obtained through tournamentForest for trials 1 to 3 are presented in **Table 2**. Disregarding group membership value, the correct classification rates for trials 1 to 3 were 50.91% (Southwestern Europe), 82.73% (Europe), and 79.07% (Europe), respectively. After the individuals whose group membership value was less than 0.80 were discarded, the correct classification rates achieved in trials 1 to 3 were 54.65% (trial 1), 86.05% (trial 2), 82.61% (trial 3). The most common classifications, other than Southwestern Europe and Europe, included Northern and Central Europe and Northeast Africa (**Table 2**).

Table 2. Classification rates for tournamentForest-AncesTrees: Trial 1-Trial 3.

Trial #	Estimated Population Affinity	% Classification (n/N)	
		No group membership values	**Group membership values ≥ 0.8**
TRIAL #1: 12-way random forest (12 ancestral groups)	Northern Asia & Arctic	0 (0/110)	0 (0/86)
	North America	0.91 (1/110)	0 (0/86)
	South America	2.73 (3/110)	2.33 (2/86)
	Northern & Central Europe	31.82 (35/110)	31.40 (27/86)
	Southwestern Europe	**50.91 (56/110)**	**54.65 (47/86)**
	Northeast Africa	5.45 (6/110)	6.98 (6/86)
	Sub-Saharan Africa	1.82 (2/110)	1.16 (1/86)
	South Asia	1.82 (2/110)	1.16 (1/86)
	East Asia	0.91 (1/110)	0 (0/86)
	Southeast Asia	0.91 (1/110)	1.16 (1/86)
	Polynesia	1.82 (2/110)	0 (0/86)
	Australia & Melanesia	0.91 (1/110)	1.16 (1/86)
Trial #2: 9-way random forest (9 ancestral groups)	Northern Asia & Arctic	0 (0/110)	0 (0/86)
	North & South America	3.64 (4/110)	2.33 (2/86)
	Europe	**82.73 (91/110)**	**86.05 (74/86)**
	Northeast Africa	5.45 (6/110)	6.98 (6/86)
	Sub-Saharan Africa	1.82 (2/110)	1.16 (1/86)
	South Asia	1.82 (2/110)	1.16 (1/86)
	East & Southeast Asia	1.82 (2/110)	1.16 (1/86)
	Polynesia	1.82 (2/110)	0 (0/86)
	Australia & Melanesia	0.91 (1/110)	1.16 (1/86)
Trial #3: 9-way random forest (9 ancestral groups)	Northern Asia & Arctic	0 (0/110)	0 (0/92)
	North & South America	0 (0/110)	0 (0/92)
	Europe	**79.09 (87/110)**	**82.61 (76/92)**
	Northeast Africa	12.73 (14/110)	10.87 (10/92)
	Sub-Saharan Africa	1.82 (2/110)	1.09 (1/92)
	South Asia	2.73 (3/110)	3.26 (3/92)
	East & Southeast Asia	0.91 (1/110)	0 (0/92)
	Polynesia	1.82 (2/110)	2.17 (2/92)
	Australia & Melanesia	0.91 (1/110)	0 (0/92)

In bold = Correct estimated group

To explore the accuracy rates differences between males and females, correct classification percentages were calculated for each sex separately (**Figure 1**). The results are evaluated both considering no membership value as well as accounting for the 0.80 threshold. For trials 1 and 2, 65.31-70.00% and 85.71-87.50% of males were correctly classified as Southwestern European and European, respectively. For their female counterparts, 40.98-41.30% and 80.33-84.87% of females were correctly classified as Southwestern European and European, respectively. The correct classification rates obtained for males through trial 3 ranged from 83.67% to 88.10%, whereas for females it ranged between 75.41% and 78.00%.

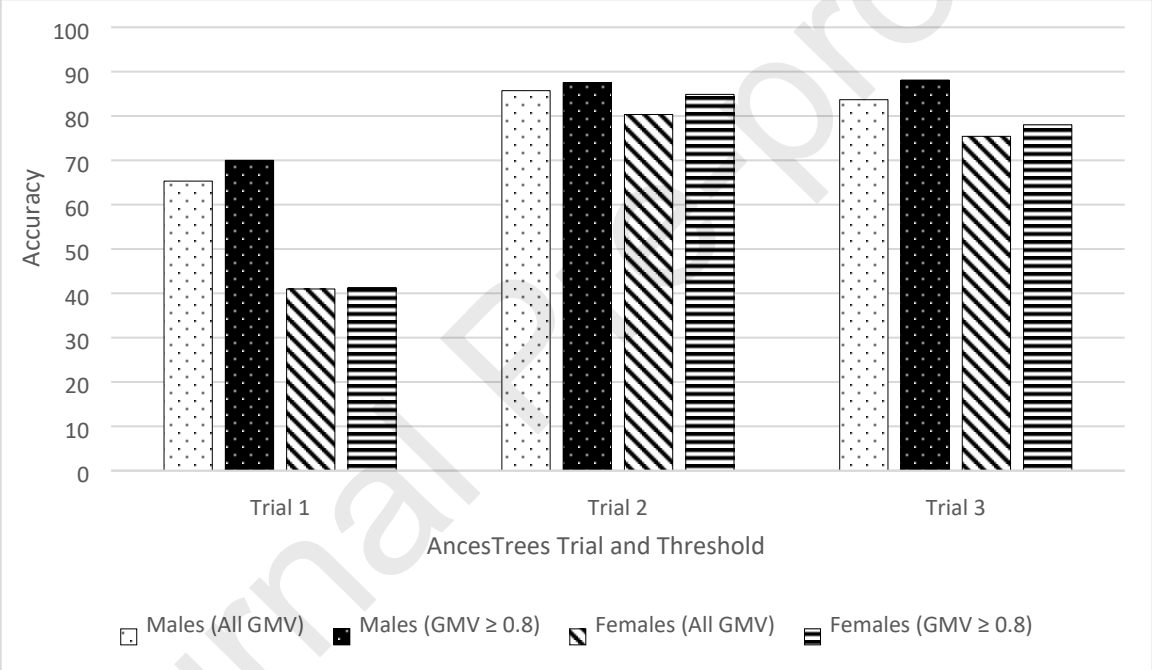

Figure 1. Accuracy rates of males and females using tournamentForest for AncesTrees trials 1-3 (GMV=group membership value).

Table 3 summarizes the ancestralForest classification rates. When group membership values were not considered, the correct classification rates for trials 4 and 5 were 34.55% (Europe) and 70% (Southwestern Europe), respectively. Employing a group membership value threshold greater than or equal to 0.80 resulted in the correct classification rate for trial 4 to drop to 0.00%, whereas it increased to 100% for trial 5. Other than the Southwestern European

and European target groups, individuals were most commonly classified as Northeast Africa, Northern and Central Europe, Northern Asia and Artic, and North and South America (**Table 3**) For simplicity, the classification rates of the remaining population groups for trial 4 have been combined.

Table 3. Classification rates for ancestralForest-AncesTrees: Trial 4 and Trial 5.

Trial #	Estimated Population Affinity	% Classification (n/N)	
		No group membership values	**Group membership values ≥ 0.8**
Trial #4: 9-way random forest (9 ancestral groups)	Northern Asia & Arctic	18.18 (20/110)	0 (0/0)
	North & South America	16.36 (18/110)	0 (0/0)
	Europe	**34.55 (38/110)**	**0 (0/0)**
	Northeast Africa	15.45 (17/110)	0 (0/0)
	Remaining groups	16.36 (18/110)	0 (0/0)
Trial #5: 12-way random forest (4 ancestral groups)	Northern & Central Europe	19.09 (21/110)	0 (0/27)
	Southwestern Europe	**70 (77/110)**	**100 (27/27)**
	Northeast Africa	10.91 (12/110)	0 (0/27)
	Sub-Saharan Africa	0 (0/110)	0 (0/27)

In bold = Correct estimated group

Regarding sex classification rates differences with no membership value and with 0.80 membership threshold, male correct classification rates ranged from 0.00% to 30.78% (trial 4) and from 79.59% to 100% (trial 5), while for females the rates ranged between 0.00% and 31.15% (trial 4) and 63.93% and 100.00% (trial 5) (**Figure 2**).

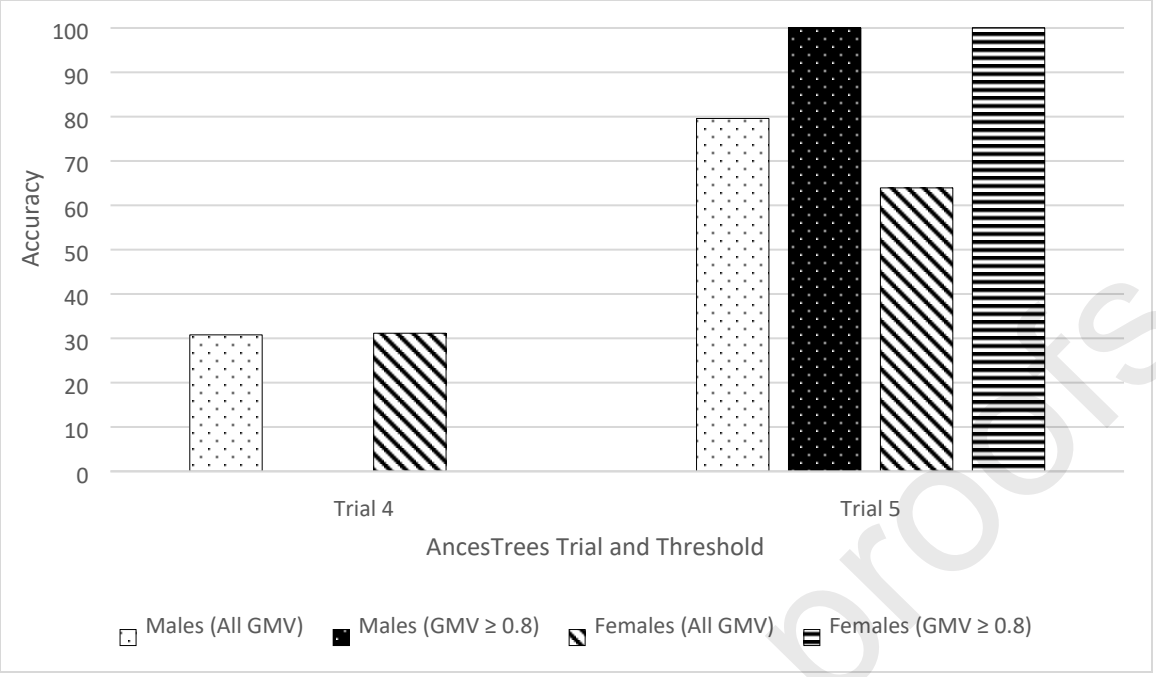

Figure 2. Accuracy rates of males and females using ancestralForest for AncesTrees trials 4 & 5 (GMV=group membership value).

3.2. (hu)MANid

Overall, out of the 114 individuals in the sample, a total of 107 individuals had the required measurements to use the (hu)MANid program.

Individuals with a PP of less than 0.50 were excluded, leaving 51 individuals for analysis. The two thresholds for PP (0.50 and 0.80) were applied and the data was analyzed first without taking into consideration correct sex estimation (**Table 4**). The results demonstrate that 70.59% of the sample was correctly classified as either WM (20c) or WF (20c) with a PP of 0.50 or more. After individuals with a PP less than 0.80 were excluded, five individuals remained. Regardless of sex, 80% were correctly classified as either WM (20c) or WF (20c). For simplicity, only the four population groups with the highest classification percentages are displayed in **Table 4**. The remaining groups have been combined.

Table 4. Classification rates for (hu)MANid, regardless of sex (posterior probability ≥ 0.5 and ≥ 0.8).

Estimated Population Affinity	% Classification (n/N)	
	Posterior Probability ≥ 0.50	Posterior Probability ≥ 0.8
CAMM or CAMF	7.84 (4/51)	0 (0/5)
GUATM or GUATF	5.88 (3/51)	0 (0/5)
KORM or KORF	11.76 (6/51)	20 (1/5)
WM (20c) or WF (20c)	**70.59 (36/51)**	**80 (4/5)**
Remaining groups	3.92 (2/51)	0 (0/0)

In bold = Correct estimated group

Table 5 presents a summary of the classification rates for the combination of accuracies for both population affinity and sex estimations. When sex was included in the assessment with a PP greater than 0.50, 62.75% were correctly estimated for both population affinity and sex, 5.88% were correctly classified for population affinity but not for sex, and 27.45% were correctly classified for sex but not for population affinity. When the PP threshold was increased to 0.80, 80.00% were correctly classified for both population affinity and sex, none were correctly classified for population affinity but not for sex, and 20% were correctly classified for sex but not for population affinity.

Table 5. Classification rates for (hu)MANid, including both population affinity and sex (posterior probability ≥ 0.5 and ≥ 0.8).

Classification	% Classification (n/N)	
	Posterior Probability ≥ 0.50	Posterior Probability ≥ 0.8
Correct population affinity, correct sex	62.75 (32/51)	80 (4/5)

4. Discussion

In the identification process, several pieces of information are needed to narrow down the identity of the unknown individual. Among them, population affinity estimation entails a

challenge due to aspects such as population sample under-representation, oversimplistic geographical classification groups or the lack of statistical analysis, among others [35, 36]. Instead of abandoning the practice, it has been suggested that the estimation of population affinity should be fully revised as it is still valuable in forensic investigations [5]. Thus, methods are being developed to improve some of the aforementioned issues, with emphasis placed on statistical analysis through the application of new mathematical approaches both for morphoscopic and metric assessments [17, 37]. To ensure that method reliability is preserved, validation studies must be conducted on the existing methods using different target samples. To date, only two other validation studies have been conducted using AncesTrees [17, 21], and three performed on testing (hu)MANid [18, 38, 39]. The present research revises the reliability of both software tools on a Spanish contemporary sample with our results indicating a wide range of classification accuracy rates depending on the model and the software used. Overall, correct classification rates increased as group membership value and posterior probability thresholds increased (except for AncesTrees trial 4); however, the reduced sample size for this analysis must be taken into consideration throughout the interpretation and discussion of the outcome (**Figure 3**).

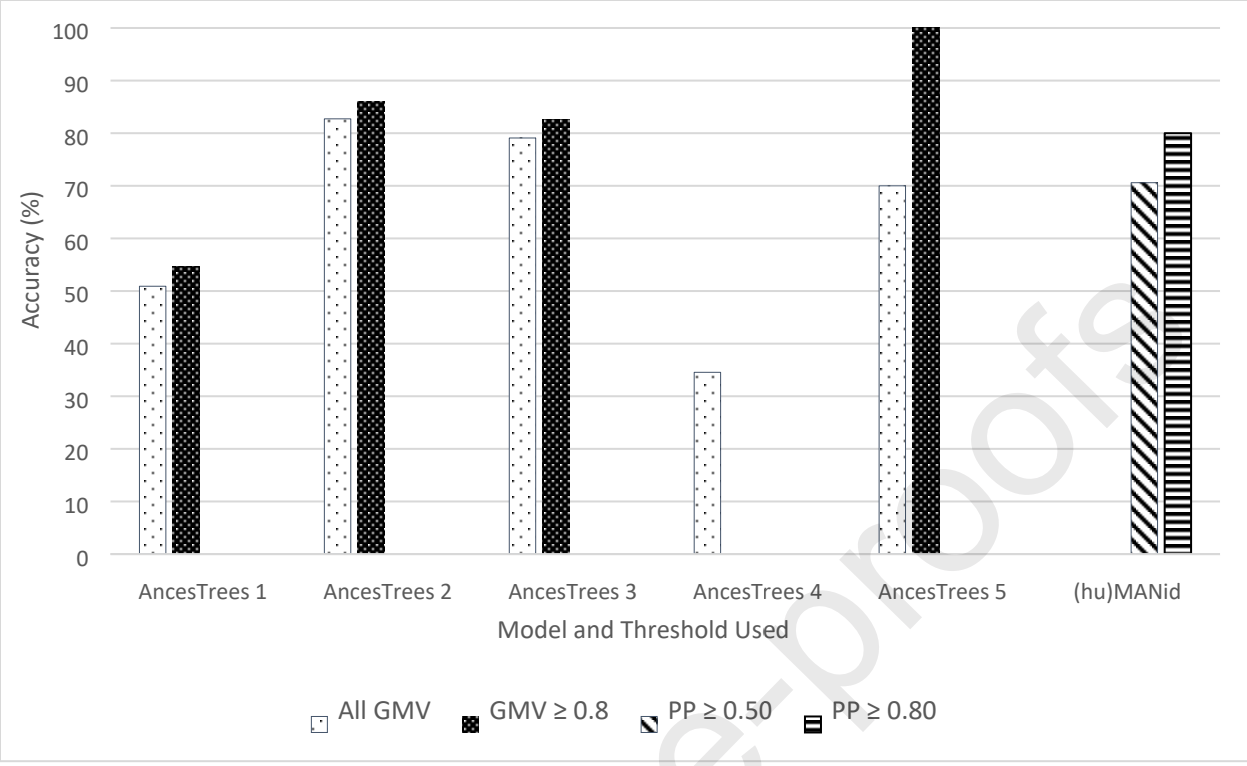

Figure 3. Comparison of accuracy rates achieved by AncesTrees and (hu)MANid (GMV=group membership value; PP=posterior probabilities).

Regarding our results, if prior information is not available, AncesTrees tournamentForest produced accuracy rates approximately 30% higher in trial 2 than in trial 1 (**Table 2**). It appears that the selection of 9 target groups instead of 12 seems to increase the accuracy rates, in accordance with other authors [17]. Trial 3 produced similar results as trial 2, suggesting that increasing the number of trees did not have a dramatic impact on the outcome. Fernandes *et al.* [21] applied the same settings to estimate population affinity on a Brazilian sample obtaining accuracy rates higher for the European-Brazilian individuals than the ones reported for the Spanish sample. Furthermore, Navega *et al.* [17] conducted two random forest trials on Portuguese and African individuals including 6 and 2 ancestral groups, respectively, with correct classification ranging from 80% to 94%. In a study on population affinity estimation for Southern European populations, tibia osteometric parameters suggested differences between Portuguese and Spanish individuals when direct comparisons were performed separately on males and females [40]. Moreover, Spanish individuals showed the lowest overall classification

rates followed by the Portuguese and Italian samples, possibly indicating some degree of overlapping for these populations. Our accuracy rates are slightly lower than those reported by Navega *et al.* [17], but higher when compared to Kranioti *et al*. [40]. These results might indicate that using AncesTrees for population affinity estimation of Spanish individuals allows for a higher classification than the tibial multinomial logistic regression formulae, although straight forward comparisons might be done with caution as the reference groups selected are less specific in the online tool (e.g., Europe or Southern Europe). Moreover, the degree of overlap due to biological affinity between the Southern samples will imply an intrinsic lower discrimination power.

The next set of classification involves ancestralForest for which the skull will be allocated based on the highest likelihood to one of the available selected groups [21]. The lowest classification accuracy was produced by trial 4 in which 34.55% of individuals were correctly classified as European, with none being correctly allocated with more than 0.80 group membership value. Conversely, 70.00-100.00% of individuals were correctly classified as Southwestern European in trial 5 when only European and African populations were pre-selected. Comparing the Spanish sample to the European-Brazilian sample [21], our classification accuracies are lower for trial 4 and similar for trial 5. The reasons behind the discrepancies and similarities between studies performed applying AncesTrees on similar Southern European populations and other geographically distinct samples might be related to gene flow, migratory forces, climate and admixture, among others [41]. Moreover, additional methodological issues such as sample size, classification strategies or individuals self-reported ancestral groups should be considered when interpreting the outcomes [17, 21, 40].

Overall, the accuracy rates achieved using AncesTrees in the present study are higher than those reported on an archaeological Spanish sample using FORDISC2 [22]. In this study, the archaeological sample was reported to be similar to the Howells' Egyptian series [27, 31, 32]. In 711 A.D., after the Umayyad had conquered the levant and reached Northeast Africa, the Islamic forces then occupied Spain for several hundred years [42, 43]. As a result, the similarity between the 16th-17th century Spanish sample and the Egyptian series may have been the product of a biological affinity to the Umayyad people [22, 44]. The Egyptian series is included

in the reference sample for AncesTrees and a percentage of individuals was allocated to the North East African group for most of our classification models (**Table 2** and **3**). Furthermore, the second membership group in the ranking was Northern and Central Europe, in accordance with Ubelaker *et al.* [22]. This could indicate the potential biological affinity reflecting the settlement patterns and various population influxes, such as the Romanization and Germanic invasions occurring in the Iberian Peninsula since approximately 800 B.C [43]. Moreover, some individuals (2.7-16%) were classified through the different models as North and South American (either separated or in combination) maybe reflecting the population influx between Spain and Latin America that started in the 15th century [45].

In relation to the power of group discrimination of the mandible in our sample, (hu)MANid correctly classified 70-80% of the sample as white, while 62.75- 80% of individuals were correctly estimated for both population affinity and sex (**Table 4**). These results are comparable and slightly higher to those achieved by Berg and Kenyhercz [18]. When compared to the accuracy rates achieved by Ubelaker *et al.* [22] using FORDISC2, those for (hu)MANid are higher when the PP is accounted for in the assessment, although sample size differences need to be acknowledged. One previous validation study using (hu)MANid on sub-adult individuals reported both correct population affinity and sex estimation rates between 18.37% and 44.83% [38], but comparisons might be made with caution due to the nature of the sample. Furthermore, the white origin group not only obtained the highest accuracy but also was more often selected than any other ancestral groups, suggesting the possibility of relatively smaller mandibles for this population in comparison with the other reference samples [38]. Regarding the classification groups obtained through our analysis, Korean, Cambodian and Guatemalan populations were next in the ranking of allocated percentages.

Apart from the aforementioned issues affecting the population allocation outcome in this study, sex is an intrinsic factor to further consider. Although sexual dimorphism is not as pronounced in humans as it is in other primates, skeletal differences exist between the sexes [46]. A study using a geometric-morphometric approach demonstrated that size differences did not have a significant effect on craniofacial shape, although sex differences accounting for size did [47]. Size differences between males and females have been reported on the Spanish

population for cranial, dental and postcranial elements [48-50]. Our results using the two online methods demonstrate an overall higher accuracy rate for males than for females, perhaps accounting for the sexual dimorphic differences. The nature of the data does not allow for more inferences, thus, future research using approaches such as geometric-morphometric as well as considering other variables such as mobility or family history will provide further insights.

Concerning practical aspects of the software online tools tested, the data entry process is straightforward, but very time-consuming for both AncesTrees [17] and (hu)MANid [18], as each parameter for each individual must be entered into the software one-by-one. AncesTrees provides a validation tab, which alerts the user when one or more of the inputted variables appears too high or low, helping the user to avoid incorrect data input. In relation to the output produced by each software, the (hu)MANid website provides an explanation of the PP and other statistical outputs assisting the user to understand the reliability of the classifications that the program produces. Including a similar section on the AncesTrees website would help in the interpretation of the significance of the results obtained. Understanding the statistical approach behind each technique is important as both software applications will force a classification into the most similar reference population in its database, even if the target parameters are not that similar to the reference sample [22].

5. Conclusion

Considering the importance of population affinity estimation in the identification process, it is critical that accurate and reliable methods are being employed. Moreover, requirements set up by *Daubert* [51] and the SWGANTH [52] call for a revision of current methodologies.

The purpose of this research was to explore the accuracy rates of two population affinity estimation software when applied to a modern Spanish sample. Our results suggest that if both the cranium and the mandible are available, the assessment should be undertaken using the cranium because overall it produced higher accuracy rates than the mandible alone. In a situation where the mandible is the only element available, PP should be considered to ensure accurate outcomes. Furthermore, the user might take into account the skeletal element under analysis, the model selected as well as the reference samples as those will have an impact on

group allocation [21]. Accounting for the different statistical approaches used by each software, group membership likelihood and PP remains crucial for the interpretation of the results.

Future research on population affinity should focus on including underrepresented populations because it can potentially improve and refine the classification accuracies. Furthermore, efforts should be made to be more consistent in the terminology used to describe biogeographic origin and to further understand population structure as well as the potential underlying factors influencing human variation.

Acknowledgement

The authors would like to thank Del Río Muñoz for making available the raw data in his PhD thesis.

References

[1] Iscan MY, Steyn M, The human skeleton in forensic medicine, Charles C Thomas Publisher, Illinois, 2013.

[2] Austin D, King RE, The biological profile of unidentified human remains in a forensic context, Academic Forensic Pathology. 6 (2016) 370–390.

[3] Ross AH, Pilloud MA, The need to incorporate human variation and evolutionary theory in forensic anthropology: A call for reform, American Journal of Physical Anthropology. 176 (2021) 672–683.

[4] Bethard JD, DiGangi EA, Letter to the Editor—Moving Beyond a Lost Cause: Forensic Anthropology and Ancestry Estimates in the United States, Journal of Forensic Sciences. 65 (2020) 1791–1792.

[5] Stull KE, Bartelink EJ, Klales AR, Berg GE, Kenyhercz MW, L'Abbé EN, Go MC, McCormick K, Mariscal C, Commentary on: 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. doi: 10.1111/1556-4029.14513., Journal of Forensic Sciences. 66 (2021) 417–420.

[6] DiBennardo R, Taylor JB, Multiple discriminant function analysis of sex and race in the post-cranial skeleton, American Journal of Physical Anthropology. 61 (1983) 305–314.

[7] Brooks S, Suchey JM, Skeletal age determination based on the os pubis: a comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods, Human Evolution. 5 (1990) 227–238.

[8] Gulhan O, Harrison K, Kiris A, A new computer-tomography-based method of sex estimation: Development of Turkish population-specific standards, Forensic Science International. 255 (2015) 2–8.

[9] Kranioti EF, Garcia-Donas JG, Can IO, Ekizoglu O, Ancestry estimation of three Mediterranean populations based on cranial metrics, Forensic Science International. 286 (2018) 265.e1-265.e8.

[10] Algee-Hewitt BFB, Coelho C, Navega D, Cunha E, 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, Academic Press, 2020.

[11] Giles E, Elliot O, Race identification from cranial measurements, Journal of Forensic Sciences. 7 (1962) 147–157.

[12] Wright RV, Correlation between cranial form and geography in Homo sapiens: CRANID–a computer program for forensic and other applications, Archaeology in Oceania. 27 (1992) 128–134.

[13] Jantz RL, Ousley SD, FORDISC, version 3.1, University of Tennessee, Knoxville, TN, 2005.

[14] Urbanova P, Kralik M, COLIPR (Coordinate-based Legal Investigation Program) [Software], Masarykova univerzita, Brno, 2008. https://is.muni.cz/publication/764216.

[15] Hefner JT, Cranial nonmetric variation and estimating ancestry, Journal of Forensic Sciences. 54 (2009) 985–995.

[16] Ousley SD, 3skull [computer program]. Windows version 1.76., 2014. https://www.statsmachine.net/software/3Skull/.

[17] Navega D, Coelho C, Vicente R, Ferreira MT, Wasterlain S, Cunha E, AncesTrees: Ancestry estimation with randomized decision trees, International Journal of Legal Medicine. 129 (2015) 1145–1153.

[18] Berg GE, Kenyhercz MW, Introducing human mandible identification [(hu)MANid]: A free, web-based GUI to classify human mandibles, Journal of Forensic Sciences. 62 (2017) 1592–1598.

[19] Urbanova P, Ross AH, Jurda M, Nogueira MI, Testing the reliability of software tools in sex and ancestry estimation in a multi-ancestral Brazilian sample, Legal Medicine. 16 (2014) 264–273.

[20] Liebenberg L, L'Abbé EN, Stull KE, Population differences in the postcrania of modern South Africans and the implications for ancestry estimation, Forensic Science International. 257 (2015) 522–529.

[21] Fernandes LCC, Bento MIC, Rabello PM, Soriano EP, Navega D, Junior ED, Cunha E, Analysis of the accuracy of AncesTrees software in ancestry estimation in Brazilian identified sample, Advances in Anthropology. 11 (2021) 163–178.

[22] Ubelaker DH, Ross AH, Graver SM, Application of Forensic Discriminant Functions to a Spanish Cranial Sample, Forensic Science Communications. 4 (2002) 1–3.

[23] Elliott M, Collard M, FORDISC and the determination of ancestry from cranial measurements, Biology Letters. 5 (2009) 849–852.

[24] Klales AR, Kenyhercz MW, Morphological assessment of ancestry using cranial macromorphoscopics, Journal of Forensic Sciences. 60 (2015) 13–20.

[25] Jacometti V, Estimativa da ancestralidade em Antropologia Forense por meio do software “AncesTrees” em medidas cranianas de uma amostra brasileira, Universidade de Sao Paulo, 2018.

[26] Del Río Muñoz PA, Estudio anthropológico-forense, anthropométrico y morfológico, de la colección de la Escuela de Medicina Legal de Madrid, Doctoral Thesis, Universidad Complutense de Madrid, 2000.

[27] Howells WW, Cranial variation in man, Peabody Museum of Archaeology and Ethnology, Harvard University, Cambridge, Massachusetts, 1973.

[28] Martin R, Saller K, Lehrbuch der Anthropologie in systematischer Darstellung, 3rd ed., Gustav Fischer Verlag, Stuttgart, 1957.

[29] Buikstra JE, Ubelaker DH, Standards for data collection from human skeletal remains, Arkansas Archeological Survey Research Series, Fayetteville, AR, 1994.

[30] DiGangi EA, Hefner JT, Ancestry estimation - Research methods in human skeletal biology, Academic Press, 2013. https://doi.org/10.1016/C2010-0-65850-0.

[31] Howells WW, Skull shapes and the map: Craniometric analyses in the dispersion of modern *homo*., Peabody Museum of Archaeology and Ethnology, Harvard University, Cambridge, Massachusetts, 1989.

[32] Howells WW, Who’s who in skulls: Ethnic identification of crania from measurements, Papers of the Peabody Museum of Archaeology and Ethnology. 82 (1995).

[33] Obertova Z, Stewart A, Cattaneo C, Statistics and probability in forensic anthropology, Academic Press, London, 2020.

[34] Berg GE, Biological affinity and sex determination using morphometric and morphoscopic variables from the human mandible, Doctoral Thesis, University of Tennessee, Knoxville, 2008.

[35] Dunn RR, Spiros MC, Kamnikar KR, Plemons AM, Hefner JT, Ancestry estimation in forensic anthropology: A review, Wiley Interdisciplinary Reviews: Forensic Science. 2 (2020) e1369.

[36] Flouri DE, Alifragki A, García-Donas JG, Kranioti EF, Ancestry estimation: advances and limitations in forensic applications, Research and Reports in Forensic Medical Science. 12 (2022) 13–24.

[37] Hefner JT, Ousley SD, Statistical classification methods for estimating ancestry using morphoscopic traits, Journal of Forensic Sciences. 59 (2014) 883–890.

[38] Tomas NR, Is the (hu)MANid Program Valid in a Pediatric Population?, Doctoral Thesis, University of Illinois at Chicago, 2020.

[39] Farhi M, Is Evaluation of the (hu) MANid Program Estimation of Sex and Ancestry in a Diverse Adult Population, Doctoral Thesis, University of Illinois at Chicago, 2022.

[40] Kranioti EF, Garcia-Donas JG, Karell MA, Cravo L, Ekizoglu O, Apostol M, Cunha E, Metric variation of the tibia in the Mediterranean: Implications in forensic identification, Forensic Science International. 299 (2019) 223–228.

[41] Botigué LR, Henn BM, Gravel S, Maples BK, Gignoux CR, Corona E, Atzmon G, Burns E, Ostrer H, Flores C, Bertranpetit J, Comas D, Bustamante CD, Gene flow from North Africa contributes to differential human genetic diversity in southern Europe, Proceedings of the National Academy of Sciences. 110 (2013) 11791–11796.

[42] Domínguez-Castro F, de Miguel JC, Vaquero JM, Gallego MC, García-Herrera R, Climatic potential of Islamic chronicles in Iberia: Extreme droughts (ad 711–1010), The Holocene. 24 (2014) 370–374.

[43] Regueiro M, Garcia-Bertrand R, Fadhlaoui-Zid K, Álvarez J, Herrera RJ, From Arabia to Iberia: A Y chromosome perspective, Gene. 564 (2015) 141–152.

[44] Ross AH, Ubelaker DH, Kimmerle EH, Implications of dimorphism, population variation, and secular change in estimating population affinity in the Iberian Peninsula, Forensic Science International. 206 (2011) 214.e1-214.e5.

[45] Salzano FM, Sans M, Interethnic admixture and the evolution of Latin American populations, Genetics and Molecular Biology. 37 (2014) 151–170.

[46] Kesterke MJ, Raffensperger ZD, Heike CL, Cunningham ML, Hecht JT, Kau CH, Nidey NL, Moreno LM, Wehby GL, Marazita ML, Weinberg SM, Using the 3D Facial Norms Database to investigate craniofacial sexual dimorphism in healthy children, adolescents, and adults, Biology of Sex Differences. 7 (2016) 1–14.

[47] Kimmerle EH, Ross A, Slice D, Sexual dimorphism in America: Geometric morphometric analysis of the craniofacial region, Journal of Forensic Sciences. 53 (2008) 54–57.

[48] Azofra-Monge A, Alemán Aguilera I, Morphometric research and sex estimation of lumbar vertebrae in a contemporary Spanish population, Forensic Science, Medicine and Pathology. 16 (2020) 216–225.

[49] Viciano J, Amores-Ampuero A, Sex estimation in a contemporary Spanish population: cranial and dental anthropometry, HOMO. 71 (2020) 189–204.

[50] Djorojevic M, Roldán C, García-Parra P, Alemán I, Botella M, Morphometric sex estimation from 3D computed tomography os coxae model and its validation in skeletal remains, International Journal of Legal Medicine. 128 (2014) 879–888.

[51] Daubert v. Merrell Dow Pharmaceuticals, 1993. https://supreme.justia.com/cases/federal/us/509/579/ (accessed February 9, 2021).

[52] Scientific Working Group for Forensic Anthropology (SWGANTH), Ancestry Assessment, (2013). https://www.nist.gov/osac/forensic-anthropology-subcommittee.

Appendix A: Supplementary Material

Descriptive statistics of skull measurements (adapted from Del Río Muñoz [26]).

Measurement	Abbreviation	*n*	Minimum (mm)	Maximum (mm)	Mean (mm)	Standard Deviation
Glabello-occipital length	GOL	108	162	199	178.82	7.679
Nasion-occipital length	NOL	107	160	196	176.83	7.256
Basion-nasion length	BNL	105	84	107	96.66	5.284
Basion-bregma height	BBH	106	113	145	129.00	6.275
Maximum cranial breadth	XCB	107	124	159	135.70	6.050
Maximum frontal breadth	XFB	104	102	130	115.58	5.969
Bizygomatic breadth	ZYB	100	108	139	122.74	6.511
Biauricular breadth	AUB	109	103	128	115.33	5.443
Biasterion breadth	ASB	107	95	121	107.21	5.905
Basion-prosthion length	BPL	99	71	101	88.60	6.222
Upper facial height	NPH	47	61	94	68.81	5.833
Nasal height	NLH	100	44	60	50.90	3.702
Bijugal breadth	JUB	101	93	117	105.69	5.390
Nasal breadth	NLB	96	19	27	22.98	2.016
Maxillo-alveolar breadth	MAB	98	43	73	54.28	5.173
Orbital height (left)	OBH	104	30	41	34.20	2.110
Orbital height (right)	OBH2	102	30	42	34.27	2.212
Orbital breadth (left)	OBB	103	33	44	37.67	2.074
Orbital breadth (right)	OBB2	101	33	45	37.87	2.086
Interorbital breadth	DKB	102	15	26	20.11	1.985

Bizygomaxillaire Anterior breadth	ZMB	99	72	99	84.36	5.224
Zygomaxillary subtense	SSS	94	17	30	22.90	2.648
Bifrontal breadth	FMB	104	82	102	92.71	4.285
Naso-frontal subtense	NAS	104	10	25	17.86	2.579
Biorbital breadth	EKB	101	83	101	92.96	4.012
Maximum cheek height	WHM	102	15	25	20.55	2.424
Frontal chord	FRC	108	98	123	109.14	5.505
Frontal subtense	FRS	107	21	33	26.28	2.289
Parietal chord	PAC	109	92	130	11.27	6.890
Parietal subtense	PAS	108	16	31	22.91	3.300
Occipital chord	OCC	108	81	109	94.43	5.001
Occipital subtense	OCS	107	21	47	29.62	4.152
Symphyseal height	GNI	53	24	41	30.85	3.754
Mandibular body height (left)	HML	104	8	36	25.72	5.562
Mandibular body height (right)	HML2	103	8	38	25.80	5.890
Mandibular body breadth (left)	TML	104	8	14	10.68	1.402
Mandibular body breadth (right)	TML2	105	8	14	10.74	1.373
Bigonial breadth	GOG	98	75	111	95.02	7.411
Bicondylar breadth	CDL	91	101	132	113.60	5.859
Maximum ramal breadth (left)	WRL	105	21	35	28.90	3.110

Maximum ramal breadth (right)	WRL2	100	21	36	29.15	3.205
Maximum ramal height	XRL	105	46	72	59.33	5.381
Mandibular length	MLT	106	57	84	70.84	5.498

Descriptive statistics of skull angle (adapted from Del Río Muñoz [26]).

Angle	**Abbreviation**	***n***	**Minimum (°)**	**Maximum (°)**	**Mean (°)**	**Standard Deviation**
Mandibular angle	MAN	106	108	148	125.04	7.533

Figures

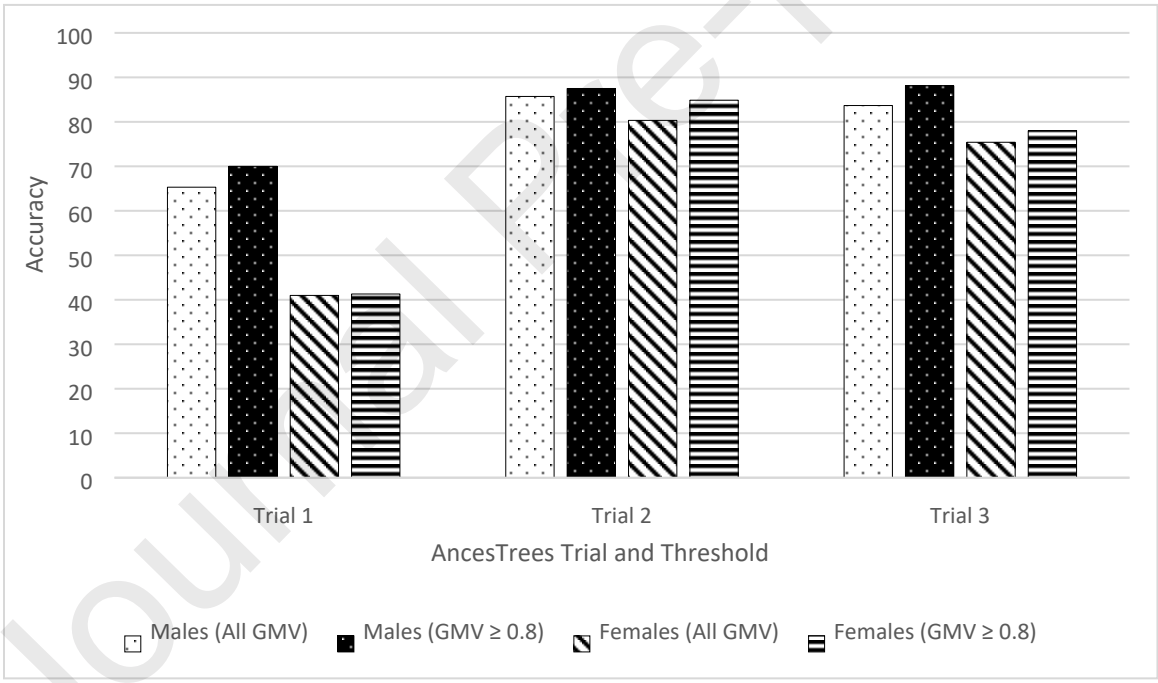

Figure 1. Accuracy rates of males and females using tournamentForest for AncesTrees trials 1-3 (GMV=group membership value).

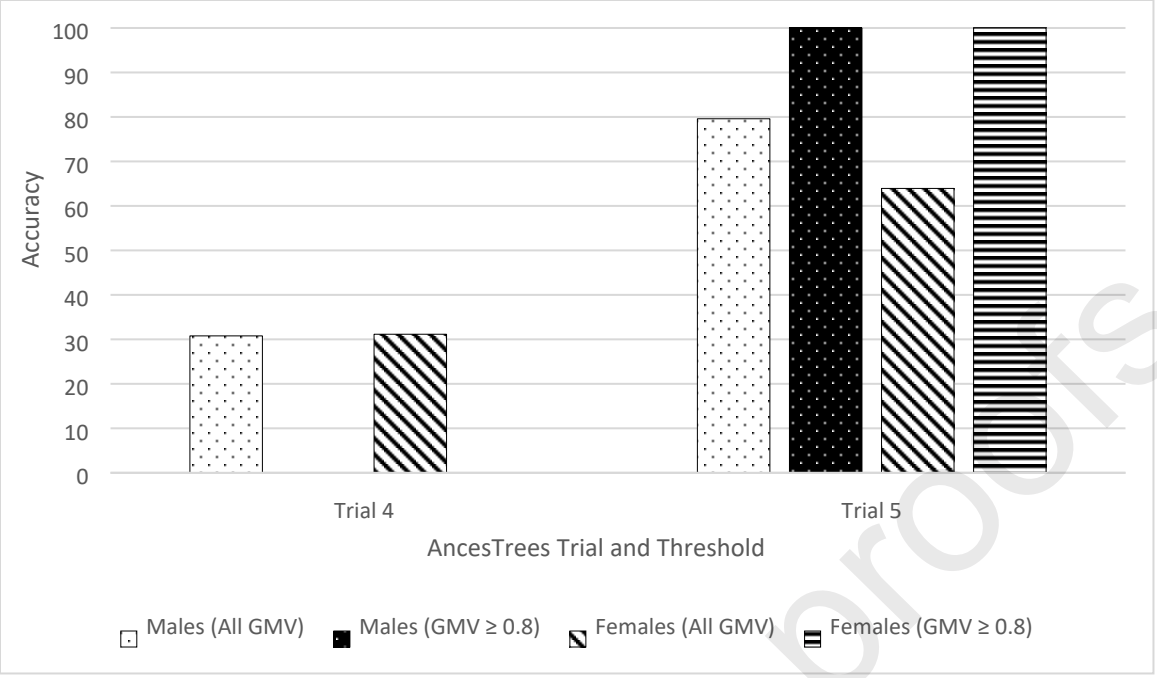

Figure 2. Accuracy rates of males and females using ancestralForest for AncesTrees trials 4 & 5 (GMV=group membership value).

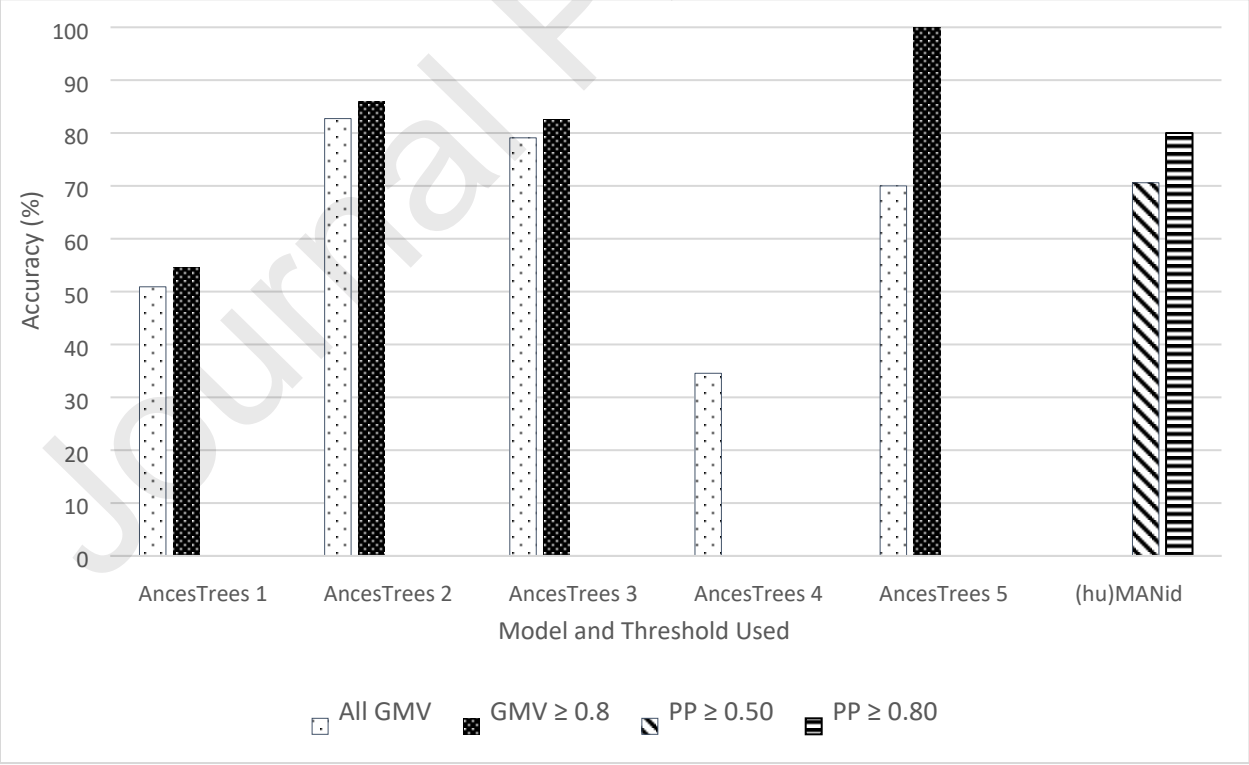

Figure 3. Comparison of accuracy rates achieved by AncesTrees and (hu)MANid (GMV=group membership value; PP=posterior probabilities).

Tables

Table 1. AncesTrees model parameters for trials 4 and 5.

Trial #	4	5
Algorithm	**ancestralForest**	**ancestralForest**
Biogeographic Ancestry Coding	*Clustered, 9 clusters*	*Clustered, 12 clusters*
Number of Trees	512	512
Number of Sub-forests	32	32
Bootstrap Fraction	63.2%	63.2%
Balanced Bootstrap	Yes	Yes
Bootstrap with Replacement	No	No
Pseudo-random Number Generator Seed	1989	1989
Parallel Computing	Yes	Yes

* All available ancestries selected.
** Northern & Central Europe, Southwestern Europe, Northeast Africa, Sub-Saharan Africa selected.

Table 2. Classification rates for tournamentForest-AncesTrees: Trial 1-Trial 3.

Trial #	Estimated Population Affinity	% Classification (n/N)	
		No group membership values	**Group membership values ≥ 0.8**
TRIAL #1: 12-way random forest (12 ancestral groups)	Northern Asia & Arctic	0 (0/110)	0 (0/86)
	North America	0.91 (1/110)	0 (0/86)
	South America	2.73 (3/110)	2.33 (2/86)
	Northern & Central Europe	31.82 (35/110)	31.40 (27/86)
	Southwestern Europe	**50.91 (56/110)**	**54.65 (47/86)**
	Northeast Africa	5.45 (6/110)	6.98 (6/86)
	Sub-Saharan Africa	1.82 (2/110)	1.16 (1/86)
	South Asia	1.82 (2/110)	1.16 (1/86)
	East Asia	0.91 (1/110)	0 (0/86)
	Southeast Asia	0.91 (1/110)	1.16 (1/86)
	Polynesia	1.82 (2/110)	0 (0/86)
	Australia & Melanesia	0.91 (1/110)	1.16 (1/86)
Trial #2: 9-way random forest (9 ancestral groups)	Northern Asia & Arctic	0 (0/110)	0 (0/86)
	North & South America	3.64 (4/110)	2.33 (2/86)
	Europe	**82.73 (91/110)**	**86.05 (74/86)**
	Northeast Africa	5.45 (6/110)	6.98 (6/86)
	Sub-Saharan Africa	1.82 (2/110)	1.16 (1/86)
	South Asia	1.82 (2/110)	1.16 (1/86)
	East & Southeast Asia	1.82 (2/110)	1.16 (1/86)
	Polynesia	1.82 (2/110)	0 (0/86)
	Australia & Melanesia	0.91 (1/110)	1.16 (1/86)
Trial #3: 9-way random forest (9 ancestral groups)	Northern Asia & Arctic	0 (0/110)	0 (0/92)
	North & South America	0 (0/110)	0 (0/92)
	Europe	**79.09 (87/110)**	**82.61 (76/92)**
	Northeast Africa	12.73 (14/110)	10.87 (10/92)
	Sub-Saharan Africa	1.82 (2/110)	1.09 (1/92)
	South Asia	2.73 (3/110)	3.26 (3/92)
	East & Southeast Asia	0.91 (1/110)	0 (0/92)
	Polynesia	1.82 (2/110)	2.17 (2/92)
	Australia & Melanesia	0.91 (1/110)	0 (0/92)

In bold = Correct estimated group

Table 3. Classification rates for ancestralForest-AncesTrees: Trial 4 and Trial 5.

Trial #	Estimated Population Affinity	% Classification (n/N)	
		No group membership values	Group membership values ≥ 0.8
Trial #4: 9-way random forest (9 ancestral groups)	Northern Asia & Arctic	18.18 (20/110)	0 (0/0)
	North & South America	16.36 (18/110)	0 (0/0)
	Europe	**34.55 (38/110)**	**0 (0/0)**
	Northeast Africa	15.45 (17/110)	0 (0/0)
	Remaining groups	16.36 (18/110)	0 (0/0)
Trial #5: 12-way random forest (4 ancestral groups)	Northern & Central Europe	19.09 (21/110)	0 (0/27)
	Southwestern Europe	**70 (77/110)**	**100 (27/27)**
	Northeast Africa	10.91 (12/110)	0 (0/27)
	Sub-Saharan Africa	0 (0/110)	0 (0/27)

In bold = Correct estimated group

Table 4. Classification rates for (hu)MANid, regardless of sex (posterior probability ≥ 0.5 and ≥ 0.8).

Estimated Population Affinity	% Classification (n/N)	
	Posterior Probability ≥ 0.50	Posterior Probability ≥ 0.8
CAMM or CAMF	7.84 (4/51)	0 (0/5)
GUATM or GUATF	5.88 (3/51)	0 (0/5)
KORM or KORF	11.76 (6/51)	20 (1/5)
WM (20c) or WF (20c)	**70.59 (36/51)**	**80 (4/5)**
Remaining groups	3.92 (2/51)	0 (0/0)

In bold = Correct estimated group

Table 5. Classification rates for (hu)MANid, including both population affinity and sex (posterior probability ≥ 0.5 and ≥ 0.8).

Classification	% Classification (n/N)	
	Posterior Probability ≥ 0.50	**Posterior Probability ≥ 0.8**
Correct population affinity, correct sex	62.75 (32/51)	80 (4/5)