ORIGINAL PAPER

Measures of divergence for binary data used in biodistance studies

Efthymia Nikita[1] · Panos Nikitas[2]

Abstract

Biodistance analysis can elucidate various aspects of past population structure. The most commonly adopted measure of divergence when estimating biodistances is the mean measure of divergence (MMD). The MMD is an unbiased estimator of population divergence but this property is lost when the dataset includes variables with very high or low frequency. In the present paper, we examine new measures of divergence based on untransformed binary data and the logit and probit transformations. It is shown that a measure of divergence based on untransformed data is a better unbiased estimator of population divergence. The conventional MMD is a satisfactory distance measure for binary data; however, it may produce biased estimations of population divergence when there are many traits with frequencies lower than 0.1 or/and greater than 0.9. Finally, the measures of divergence based on the probit and logit transformations are usually biased estimators.

Keywords Biodistance · Divergence · Transformation · Binary variables

Introduction

Biodistance analysis examines the relatedness or distance of past populations, employing skeletal and dental phenotypic data. These phenotypic data include metrics and nonmetric traits, and are used as a proxy for the genotype with the underlying assumption that phenotypic variability expresses phylogenetic variation (Relethford 2016). The observed biodistances may be attributed to various factors such as gene flow, shared ancestry, or other processes (Buikstra et al. 1990; Hefner et al. 2016; Pietrusewsky 2013); therefore, this type of analysis can elucidate important aspects of past population history.

Biodistance studies were among the first domains of physical anthropology developed as early as the eighteenth century (Blumenbach 1775). Early studies were heavily focused on craniometrics and had a typological and descriptive character (e.g., Broca 1863; review by Stojanowski 2019). From the 1960s and the emergence of New Archaeology, biodistance analysis diversified the topics addressed and encompassed various issues of past population structure and short-distance mobility along with large-scale migrations (Buikstra 1979; Corruccini 1972; Rightmire 1970). The generalization of the use of advanced non-destructive methods for data collection (e.g., geometric morphometrics), along with the enhanced availability of software to implement complex statistical tests, has contributed to biodistance studies still being broadly used in the twenty-first century (Pilloud and Hefner 2016).

Any measure used to estimate biodistances should have two main properties: (a) it should be an unbiased estimator of population divergence because even though we are estimating paired biodistances between samples, what we are actually interested in is the distance among the populations from which these samples derive, (b) it should provide the means to evaluate if the biodistances estimated are statistically significant or not.

Many measures of dissimilarity have been proposed for binary data, including Jaccard distance, simple matching distance, and Hamming distance. In biodistance studies, the most common measure used, mainly because it is an unbiased estimator of population divergence, is the mean measure of divergence (MMD). The MMD was devised by the British statistician C.A.B. Smith in order to be used by M.S. Grewal (1962) in his estimation of biological divergence across generations of laboratory mice in sublines of the C57BL strain.

✉ Efthymia Nikita
efi.nikita@gmail.com

Panos Nikitas
nikitas@chem.auth.gr

1 Science and Technology in Archaeology and Culture Research Center, The Cyprus Institute, 2121 Aglantzia, Nicosia, Cyprus

2 Department of Chemistry, Aristotle University of Thessaloniki, University Campus, 54124 Thessaloniki, Greece

Subsequently, the use of the MMD was generalized in anthropology through the work by A.C. Berry and R.J. Berry, who examined the biodistance of different human groups using nonmetric skeletal traits (e.g., Berry and Berry 1967; Berry 1974). This measure is still extensively adopted for biodistance estimation using skeletal and dental dichotomous nonmetric traits in studies across the world (e.g., Irish 2016; Nikita et al. 2012; Ullinger et al. 2005). Its use is so generalized that it has led a number of scholars to propose ways of automating its calculation (Bertsatos and Chovalopoulou 2016; Nikita 2017; Santos 2018; Sołtysiak 2011), whereas recently, a new version of a parametric bootstrap for the mean measure of divergence has been developed to make it suitable for samples with scarcity of data (Zertuche and Meza-Peñaloza 2020).

Despite its broad applicability, the MMD has two limitations: (a) it is based on the arcsine transformation, so when there are many traits with low or high frequency (less than 0.1 or greater than 0.9), it ceases being an unbiased estimator of population divergence, (b) when traits are inter-correlated, the relationships used to determine the statistical significance of the distances may not be accurate. For these reasons, datasets are usually edited prior to the use of the MMD in order to remove traits with low or high frequency as well as inter-correlated traits (Harris and Sjøvold 2004; Irish 2010 and references therein). Nonetheless, such editing may lead to the loss of important information as it often results in a substantial number of variables/nonmetric traits being eliminated from the estimation of the biodistances.

In the present paper, we examine three new measures of divergence based on untransformed data and the logit and probit transformations, additionally to the MMD. To test the performance of these measures of divergence, we use simulated data and real datasets of dental nonmetric traits. The aim is to test under what conditions the aforementioned measures of divergence are unbiased estimators of population divergence and consequently whether untransformed data or some transformation, which does not require data editing/elimination, produces better results.

Measures of divergence

A general expression

Consider two samples, 1 and 2, originating from two populations with binary data that code the presence or absence of a specific trait. If n_1, n_2 are the sample sizes and k_1, k_2 the number of individuals in the samples possessing the particular trait, then an obvious measure of divergence is the squared Euclidean distance:

$$ED = (\varphi_1 - \varphi_2)^2 \tag{1}$$

where $\varphi_1 = k_1/n_1$, $\varphi_2 = k_2/n_2$ are the proportions of the individuals in the samples possessing the trait under examination. Despite its simplicity, Eq. (1) is not an appropriate measure of divergence for biodistance studies because this distance is not an unbiased estimator of population divergence. This is because the expected value of ED, E[ED], is given by (Souza and Houghton 1977):

$$E[\text{ED}] = E\left[(\varphi_1 - \varphi_2)^2\right] = (E[(\varphi_1 - \varphi_2)])^2 + \text{Var}(\varphi_1 - \varphi_2)$$
$$= (p_1 - p_2)^2 + \text{Var}(\varphi_1) + \text{Var}(\varphi_2) \tag{2}$$

where p_1, p_2 are the expected values of φ_1, φ_2, that is, p_1, p_2 are the proportions of individuals in the two populations having the trait in question. It is seen that the expected value of ED is not equal to the corresponding Euclidean distance between the populations, $(p_1 - p_2)^2$, due to the variances $\text{Var}(\varphi_1)$, $\text{Var}(\varphi_2)$.

A direct way to define an unbiased estimator of population divergence has been adopted in developing the well-known mean measure of divergence (MMD) (Berry and Berry 1967; Berry 1974, 1976; Grewal 1962; Sjøvold 1973, 1977; Souza and Houghton 1977). In general, to define an unbiased estimator of population divergence, a data transformation is selected, $t = f(\varphi)$ that either stabilizes the variance or the variance is expressed in terms of p_i, n_i, i.e., $\text{Var}(t_i) = V_i = V(n_i)$ or $\text{Var}(t_i) = V_i = V(p_i, n_i)$, where $V(n_i)$ denotes a function of n_i and $V(p_i, n_i)$ denotes a function of p_i, n_i. Then, an unbiased estimator of population divergence is defined from

$$\text{MD} = (t_1 - t_2)^2 - (V_1 + V_2) \tag{3}$$

since, according to Eq. (2), we have:

$$E[\text{MD}] = (T_1 - T_2)^2 + V_1 + V_2 - (V_1 + V_2) = (T_1 - T_2)^2 \tag{4}$$

where T_1, T_2 are the expected values of t_1, t_2. This definition is straightforwardly extended to the case where each population consists of r traits. Then, for each trait i, we have:

$$\text{MD}_i = (t_{1i} - t_{2i})^2 - (V_{1i} + V_{2i}) \tag{5}$$

and therefore, a total measure of divergence may be defined from:

$$\text{TMD} = \sum_{i=1}^{r} \text{MD}_i = \sum_{i=1}^{r} \left\{ (t_{1i} - t_{2i})^2 - (V_{1i} + V_{2i}) \right\} \tag{6}$$

whereas the mean measure of divergence (including the commonly adopted C.A.B. Smith's MMD and its extensions) is given by:

$$MMD = TMD/r \qquad (7)$$

If the transformed data is nearly normally distributed and under the hypothesis that there is no population divergence, the quantity:

$$S = \sum_{i=1}^{r} \frac{(t_{1i} - t_{2i})^2}{V_{1i} + V_{2i}} \qquad (8)$$

will be distributed approximately as the chi-squared distribution with r degrees of freedom, χ_r^2, provided that all r traits are independent (Souza and Houghton 1977). Therefore, this statistic can be used to test the null hypothesis that TMD = 0. Alternatively, we may use the following test statistic (Sjøvold 1977; Souza and Houghton 1977):

$$T = TMD \Big/ \sqrt{2 \sum_{i=1}^{r} (V_{1i} + V_{2i})^2} \qquad (9)$$

which follows the standard normal distribution $N(0,1)$.

Data transformations

Several data transformations have been proposed for proportions, which exhibit, at least approximately, the requested properties for the variance. The most common transformation is the arcsine transformation, which is a variance-stabilizing transformation. Within the context of the mean measure of divergence, this transformation can be performed by means of Smith's formula, $t = \sin^{-1}(1 - 2\varphi)$, where $\sin^{-1}$ is the inverse trigonometric sine function, although two other transformations, the Freeman and Tukey (1950) and Anscombe (1948) transformations, perform better and have replaced Smith's formula. Using the arcsine transformation and irrespective of the specific formula used to implement this transformation, the variability of the transformed data is given by Bartlett (1947) and Sjøvold (1977):

$$Var(t) \approx 1/n \qquad (10)$$

except for very small or high φ values and low n values, lower than 20. This expression of Var(t), in combination with Eqs. (6) and (7), yields directly Smith's mean measure of divergence (MMD) (Berry and Berry 1967; Berry 1974, 1976; Grewal 1962; Harris and Sjøvold 2004; Irish 2010; Nikita 2015; Sjøvold 1973, 1977).

Apart from the arcsine transformation, there are at least three other approaches that should be taken into account: the untransformed proportions, the logit, and the probit transformation. For untransformed proportions, we have (Sjøvold 1977):

$$E[\varphi] = p \text{ and } Var(\varphi) = p(1-p)/n \qquad (11)$$

whereas for the logit transformation, $t = \text{logit}(\varphi) = \ln(\varphi/(1-\varphi))$, Bartlett (1947) suggests:

$$Var(t) \approx \frac{1}{p(1-p)n} \qquad (12)$$

For the probit transformation, it is not known what the variance expression in terms of p and n is (Bartlett 1947). However, there is a close relationship between logit and probit functions that may be expressed as (https://en.wikipedia.org/wiki/Probit):

$$\text{logit}(\varphi) \approx \text{probit}(\varphi) / \sqrt{\pi/8} \qquad (13)$$

which yields:

$$Var(\text{probit}(\varphi)) \approx Var\left(\sqrt{\frac{\pi}{8}} \, \text{logit}(\varphi) \right)$$
$$= \frac{\pi}{8} \, Var(\text{logit}(\varphi)) \approx \frac{\pi/8}{p(1-p)n} \qquad (14)$$

Equations (12) and (14) may be written as

$$Var(t) \approx \frac{c_{P,L}}{p(1-p)n} \qquad (15)$$

where $c_P = 1$ for the logit transformation and $c_L = \pi/8$ for the probit transformation. Note that Var(t) in Eq. (15) varies linearly with $1/n$ at a constant p value, whereas at a constant n value, it varies linearly with $1/p(1-p)$. Therefore, a plausible modification of Eq. (15) that may extend its validity is the following:

$$Var(t) \approx \frac{\alpha_{P,L}/p(1-p) + b_{P,L}}{n} \qquad (16)$$

where $\alpha_{P,L}$, $b_{P,L}$ are constants. When $b_{P,L} = 0$ and $\alpha_{P,L} = c_{P,L}$, Eq. (16) is reduced to Eq. (15). However, we should point out that Eq. (15) and, therefore, Eq. (16) are approximate expressions of Var(t) and, therefore, we should use simulations to determine the optimum values of $\alpha_{P,L}$, $b_{P,L}$ and $c_{P,L}$ that make these expressions of Var(t) useful in the widest possible range of p, n values. This issue is examined below in the "Materials and methods" and the "Results and discussion" sections.

Measures of divergence

Based on the results presented above for the variance of transformed/untransformed data and the general expression for the total measure of divergence, Eq. (6), we obtain the following measures of divergence:

Untransformed data

Using raw data, $t = \varphi$, and Eqs. (6) and (10) readily yield:

$$\text{UMD} = \sum_{i=1}^{r} \left\{ (\varphi_{1i} - \varphi_{2i})^2 - \frac{p_{1i}(1-p_{1i})}{n_{1i}} - \frac{p_{2i}(1-p_{2i})}{n_{2i}} \right\} \quad (17)$$

where n_{1i} is the number of individuals from sample 1 in which the presence of trait i is examined and n_{2i} is the corresponding number of individuals from sample 2.

Arcsine transformation

It results in the well-known mean measure of divergence (MMD):

$$\text{MMD} = \frac{\text{TMD}}{r} = \frac{1}{r} \sum_{i=1}^{r} \left\{ (t_{1i} - t_{2i})^2 - \frac{1}{n_{1i}} - \frac{1}{n_{2i}} \right\} \quad (18)$$

where t_{1i} and t_{2i} denote the transformed φ values of each trait per sample. As mentioned above, initially, the transformation was performed by means of Smith's formula, but now, due to better performance, the transformations suggested by Freeman-Tukey (1950) and Anscombe (1948) are preferred. The latter may be expressed as:

$$t = \sin^{-1} \left(1 - 2 \frac{k + 3/8}{n + 3/4} \right) \quad (19)$$

where k shows how many times each trait appears in a sample of n size, that is, $\varphi = k/n$. Note that for the MMD, the addition of the value 0.5 in the two last denominators of Eq. (18) has been suggested (Green and Suchey 1976; Irish 2010).

Probit transformation

The probit transformation gives:

$$\text{PMD} = \sum_{i=1}^{r} \left\{ (\text{probit}(\varphi_{1i}) - \text{probit}(\varphi_{2i}))^2 - \frac{P_{1i}}{n_{1i}} - \frac{P_{2i}}{n_{2i}} \right\} \quad (20)$$

where

$$P_{1,2} = \frac{a_P}{p_{1,2}(1-p_{1,2})} + b_P \approx \frac{c_P}{p_{1,2}(1-p_{1,2})} \approx \frac{\pi/8}{p_{1,2}(1-p_{1,2})} \quad (21)$$

Logit transformation

The logit transformation gives:

$$\text{LMD} = \sum_{i=1}^{r} \left\{ (\text{logit}(\varphi_{1i}) - \text{logit}(\varphi_{2i}))^2 - \frac{L_{1i}}{n_{1i}} - \frac{L_{2i}}{n_{2i}} \right\} \quad (22)$$

where

$$L_{1,2} = \frac{a_L}{p_{1,2}(1-p_{1,2})} + b_L \approx \frac{c_L}{p_{1,2}(1-p_{1,2})} \approx \frac{1}{p_{1,2}(1-p_{1,2})} \quad (23)$$

In all the above relationships, with the exception of the MMD, Equation (18), the population probabilities p_1, p_2 may be estimated from the corresponding sample quantities, φ_1, φ_2 since, according to Equation (11), $E[\varphi] = p$, i.e., φ is an unbiased estimator of p. However, when p is close to 0 or 1 and the sample size is relatively small, φ may be equal to 0 or 1. In this case, Bartlett's correction may be used (Harris and Sjøvold 2004), which entails the replacement of $p=0$ with $p=1/4n$, and $p=1$ with $p=1-1/4n$. Finally, although $P_{1,2}$ and $L_{1,2}$ may be approximated via Eqs. (21) and (23), i.e., $P_{1,2} = \pi/(8p_{1,2}(1-p_{1,2}))$ and $L_{1,2} = 1/(p_{1,2}(1-p_{1,2}))$, it is better to use the expressions with the constants a_P, b_P, a_L, b_L. These constants are estimated using simulations, as discussed below.

Materials and methods

Tests for data transformations

To examine the performance of the data transformations, we used a simple simulation, which estimates the conditions under which the variance of the transformed data may be expressed as a function of p, n, what this expression is under logit and probit transformation, and whether the transformed data follow a nearly normal distribution or not. The *rbinom* function of the R language, which can model individual Bernoulli trials, was used for its implementation. Based on *rbinom*, a function was written that generates random numbers from the binomial distribution, it creates samples of random numbers of various size, n, and binomial probabilities, p, it calculates the φ value of each sample of size n, and, based on the φ values, it estimates the variances after the transformations under study using 100000 iterations (samples). The function also provides histograms and Q-Q plots to test the normality of the transformed/untransformed data. The Q-Q plots were created by means of the *qqPlot* function of the *car* library. For the logit and probit transformations, Bartlett's correction was used when a φ value equal to 0 or 1 was generated (Bartlett 1936; Harris and Sjøvold 2004). Concerning constants a, b, and c in Eqs. (21) and (23), these have been estimated in Excel spreadsheets using Solver by minimizing the value of the sum of squared errors, where errors are the

differences between calculated and simulated variances. The calculations were performed in the region $0.15 \leq p \leq 0.85$ and $n \geq 50$ (Supporting Data S1). Note that even though we used $n \geq 50$ when estimating constants a, b, and c in the Solver, we have found that the same values may be used for smaller n values ($n \geq 30$).

Generation of simulated data

To test the measures of divergence presented above, we used simulated data and real datasets of dental nonmetric traits. The simulated data were multivariate binary data. In all artificial datasets, the number of samples was equal to 5 consisting of either 10, 20, 30, 40, and 50 cases or 100, 200, 300, 400, and 500 cases to test the effect of sample size. The number of traits per sample was $r = 20$. The binary data were generated given marginal distributions and a correlation matrix using the *generate.binary* function of the *MultiOrd* package of the R library. We used two types of correlation matrix: the unit correlation matrix and the matrix with the greatest correlations that is compatible with the marginal probabilities adopted for the generation of the various artificial datasets. In what concerns the marginal probabilities, p, the first sample is generated from uniformly distributed p values between $p1$ and $p2$, the second sample is generated from uniformly distributed p-values between $p1$+step and $p2$+step, the third consists of uniform values between $p1$+2*step and $p2$+2*step, and so on, where step takes the values 0, 0.005, 0.05, and 0.1, $p1$ takes the values 0, 0.05, and 0.1, and $p2$ the values 0.1, 0.15, and 0.2. Typical matrices of marginal probabilities and correlation matrices used in the present study are shown in the Supporting Data S2.

To examine the bias of the distances under study, very large samples of 10000 cases each were used to approximate populations from which 1000 samples of 10, 20, 30, 40, and 50 cases or 100, 200, 300, 400, and 500 cases were randomly drawn; their distances were computed and averaged. The bias is evaluated from the difference between population distances and averaged sample distances. Note that the distance between populations 1 and 2 is calculated from the sum of terms $(T_{1i} - T_{2i})^2$, where i ranges from 1 to r and T may be estimated either from the value of the transformation $t = f(\varphi)$ at the population, $T = f(p) \approx f(\varphi)$, or from the expected value of t, $T = E[t]$. Considering Equation (4), the latter should be preferred.

Simulation of a real dataset

The simulation of a real dataset of samples with binary data is straightforward using correlated multivariate binary variates of known marginal probabilities. However, to test whether a given measure of divergence behaves as an unbiased population estimator, we need to simulate the populations from which the dataset originates. This may be achieved if we take into account that the sample φ value is an unbiased estimator of the corresponding population p value. For this reason, in the present study, populations of size 50000 were created using as marginal probabilities the φ values of the original samples. Then, 1000 samples were randomly drawn from the populations; their distances were calculated and averaged.

When a measure of divergence is an unbiased estimator of population divergence, population and averaged sample distances coincide. Therefore, a simple scatterplot of the population and averaged sample distances of all pairs of populations/samples under study is a simple way to visualize biases and detect their magnitude.

At this point, we should clarify the following. When we simulate a population based on the sample φ value, then the corresponding population p value is practically equal to φ, $p \approx \varphi$. Similarly, for the population T value, we have $T \approx t$, where t is a transformation of φ. Therefore, according to Eqs. (3) and (6), the distance between two simulated populations i, j is greater than that between the corresponding original samples and this difference is equal to the sum of sample variances $V_i + V_j$. Consequently, if we consider a scatterplot with pairwise sample, population, and averaged sample distances, the original distances do not coincide with the population distances unless the sum $V_i + V_j$ is very small, i.e., when the sample sizes are large. The original distances coincide with the population distances under all circumstances if we subtract from each distance between populations and from each averaged distance between samples drawn randomly from the populations the sum $V_i + V_j$ estimated from the sample properties. Note that since the bias is estimated only from the difference between population and averaged sample distances and during this correction, the same quantity $(V_i + V_j)$ is subtracted from the population and averaged sample distances, the biases in the estimation of population divergence remain the same. Therefore, under this correction, if a measure of divergence is an unbiased estimator, the averaged sample distances will coincide with the distances between the original samples and between the simulated populations; otherwise, the averaged sample distances will deviate from the corresponding original and population distances.

Datasets of dental nonmetric traits

The dataset of actual archeological material of dental nonmetric traits was collected by the first author from seven archeological sites across Greece. The archeological material intentionally comprised datasets with diverse temporal and geographic distribution in order to evaluate how meaningful the obtained biodistances are. In particular, we used three Cretan assemblages: Kephala Petras rock shelter, Livari-Skiadi rock shelter, and Eleutherna, and four mainland Greek assemblages: Xeropigado, Akraiphnio, Christianoi, and Pella. The number of dental traits in the original database

was 28 (GR-28). A subset of this dataset with 12 traits (GR-12) was also examined after data editing to remove traits with $\varphi \leq 0.01$.

The Kephala Petras rock shelter lies 1 km east of the modern town of Siteia in north-eastern Crete and dates to the Early Minoan IB - Middle Minoan IB/IIA (ca. 2900-1875/1850BC) (Tsipopoulou 2010). The Livari-Skiadi rock shelter is located in south-eastern Crete and it is largely contemporary to the Kephala Petras material, dating to the Early Minoan IB - Early Minoan III (ca. 2900-2100/2050BC) (Papadatos and Sofianou 2015). The material from Eleutherna was excavated in a proto-Byzantine Basilica and dates to the 6th–7th c. AD (Themelis 1994–1996). Xeropigado is located in northwestern Greece (near the modern city of Kozani) and dates to the Early and Middle Bronze Age (2420–1730 BC) (Maniatis and Ziota 2011). Pella is in north-central Greece and dates to the Early Bronze Age (Akamatis 2009). The cemetery of ancient Akraiphnio is located at the eastern coast of Lake Kopais in central Greece (close to the modern city of Thebes) and temporally spans the Late Geometric to the Late Roman times (7th c. BC–6th c. AD) (Sabetai 1995; Vlachogianni 1997). The skeletal assemblage from Christianoi was found inside the Church of Transfiguration of the Saviour at Christianoi, in the district of Triphylia in Messene in the Peloponnese, and dates approximately to the 13th c. AD (Kappas and Sakkari 2012). The bioarcheological analysis of the above assemblages has been published in Nikita et al. (2019) for Akraiphnio, Kalliga (2015) for Pella, Triantaphyllou (2010) for Xeropigago, Triantaphyllou (2012) for Kephala Petras, Bourbou (2004) for Eleutherna, and Triantaphyllou (2016) for Livari-Skiadi, while the material from Christianoi is currently under preparation for publication.

The main feature of the dental dataset is the great percentage of traits with small and very small φ values (Table 1). For more details about this dataset, see the Supporting Data S3.

Software to test distance measures

The generation of artificial data for distance measures, as well as the calculation of the distance measures presented in this study and their p values, was carried out by means of four homemade functions in R: *simbiasbinMDs*, *simperbinMDs*, *biasbinMDs*, and *perbinMDs*.

Table 1 Percentage of low and high trait frequencies in datasets of dental nonmetric traits from Greek archeological assemblages

Dataset	Traits	$\varphi \leq 0.1$	$\varphi \leq 0.05$	$\varphi \leq 0.01$	$\varphi \geq 0.9$
GR-28	28	51.0%	35.7%	20.9%	0
GR-12	12	8.3%	3.6%	0	0

The *simbiasbinMDs* function is used to generate artificial data for five samples of 20 traits based on certain marginal probabilities and correlation matrices. Based on these data, pairwise distances between populations are calculated, where each population is approximated by a sample of 10000 cases. The obtained distances are compared to pairwise distances between samples randomly drawn from the corresponding populations. The latter distances are averaged over 1000 samples each. The function provides also dendrograms for sample and averaged distances using the hclust() function.

The *simperbinMDs* function is used to estimate the accuracy of the p values of the estimated distances obtained from artificial data. For this estimation, p values obtained for each distance from the two test statistics, S and T, Eqs. (8) and (9), are compared to p values estimated from the permutation method. The permutation method adopted involves the following steps: (1) calculation of the value d of a certain measure of divergence and the values S and T of the test statistics using the original dataset, (2) random redistribution of all cases into new groups with sizes equal to the original ones, (3) calculation of d_i and the values S_i and T_i of the test statistics based on the new groups, (4) repetition of steps 2 and 3 at least $M=1000$ times, and (5) estimation of three p values for each distance from the ratios $N(d_i \geq d)/M$, $N(S_i \geq S)/M$, and $N(T_i \geq T)/M$, where $N(x \geq y)$ is the number of cases where x is greater than or equal to y.

The *biasbinMDs* function is used to assess biases in the estimation of population divergence. The original dataset is used to simulate the respective populations and calculate all pairwise population distances using the measures of divergence discussed in the present study. Then, these distances are compared to the corresponding pairwise distances between samples randomly drawn from the populations. As in the first function, the latter distances are averaged over 1000 samples each. From the comparison, we can assess which of the distances calculated on the dataset under study are unbiased estimators of population divergence. The function provides also dendrograms of the distances of the original dataset and the averaged distances.

Finally, the *perbinMDs* function estimates the accuracy of p values of distances calculated on a specific dataset. It works like the *simperbinMDs* function.

All software material is given as Supporting Data S4.

Results and discussion

Results for data transformations

Some of the obtained results are shown in Table 2 and Fig. 1, whereas more details are presented in the Supporting Data S1. The results show that, apart from the untransformed data, where the relationship $Var(\varphi) = p(1-p)/n$ is valid for every

Table 2 Summary of data transformations for proportions

Transformation	Variance	Range
No transformation	$p(1-p)/n$	Every p, n
Arcsine	$1/n$	$0.1 \leq p \leq 0.9$ and $n \geq 20$
Probit	$\frac{\frac{0.26}{p(1-p)}+0.57}{n} \approx$ $\frac{0.36}{p(1-p)n} \approx \frac{\pi/8}{p(1-p)n}$	$0.15 \leq p \leq 0.85$ and $n \geq 30$
Logit	$\frac{\frac{1.31}{p(1-p)}-1.17}{n} \approx$ $\frac{1.1}{p(1-p)n} \approx \frac{1}{p(1-p)n}$	$0.15 \leq p \leq 0.85$ and $n \geq 30$

p and n value, all other expressions exhibit, as expected, asymptotic variance stabilization as well as asymptotic normality. Asymptotic normality is also shown from the untransformed data.

For the variance of the transformed data, the arcsine transformation performs better since it fails for $p \leq 0.1$ or $p \geq 0.9$ when $n \leq 20$, whereas for probit/logit, these limits are $p \leq 0.15$ or $p \geq 0.85$ when $n \leq 30$. In what concerns the normality of the φ, t data, in general, there are rather strong deviations from normality when $n \leq 50$, irrespective of the p value. The normality is improved as n increases, especially when $p > 0.1$ or $p < 0.9$. This holds for the untransformed data and the arcsine transformation, whereas data obtained from probit/logit transformations exhibit more pronounced deviations from normality especially when $p \leq 0.1$ or $p \geq 0.9$.

The asymptotic nature of variance stabilization and data normality is expected to affect the bias of the calculated distances and the validity of the significance tests. The extent of this effect can only be assessed through extensive analysis of artificial and real data, as attempted in the current study.

The problem of negative distances

A main issue when using the MMD is the calculation of negative MMDs when the compared samples come from populations that are biologically very similar. In this case, $t_{1i} \approx t_{2i}$ and a negative MMD may be calculated from the terms $-1/n_{1i}$ and $-1/n_{2i}$. Because negative distances are considered to be biologically meaningless, negative MMD values are changed to zero prior to subsequent multivariate analyses (Harris and Sjøvold 2004) or a proper constant is added to all pairwise MMDs (Ossenberg et al. 2006). An alternative suggestion is to delete traits that have a negative contribution to the MMD (Harris and Sjøvold 2004).

If negative MMD values are considered biologically meaningless and should be transformed to 0, the same approach should be followed for the individual traits that have a negative contribution in the calculation of the MMD. However, if we transform the negative contributions to 0, the MMD is no longer an unbiased estimator of population divergence. This is

because the calculated sample distances increase since negative contributions become 0, whereas this procedure leaves population distances unaffected because in the population the size n is very large and, therefore, $1/n$ tends to zero, resulting in non-negative contributions to the population MMDs. Consequently, transforming negative MMD contributions to zero should be avoided.

The suggestion to delete traits that have a negative contribution to the MMD seems to solve the above problem; however, it should also be avoided for the following reason. The elimination of traits that correspond to biologically similar populations enhances the differences between the populations, which may result in false conclusions about their biodistance. Moreover, for other pairs of samples, these traits may correspond to biologically distant populations. Therefore, the elimination of traits may exclude negative values in the MMD, but it does not exclude the possibility of a strong bias of the MMDs in relation to population biodistances.

Finally, Ossenberg's et al. (2006) suggestion is mathematically equivalent to adding a constant (the same constant) to each trait contribution, i.e., $(t_{1i} - t_{2i})^2 - 1/n_{1i} - 1/n_{2i} + const$, where $const$ is the absolute value of the largest negative value of all $(t_{1i} - t_{2i})^2 - 1/n_{1i} - 1/n_{2i}$ values over all pairs of samples. Under this transformation, all the MMDs are raised by the amount $r*const$, and for this reason, no negative MMDs are computed. Moreover, the MMD remains an unbiased estimator of population divergence. Thus, consider that we create a dataset of populations using correlated multivariate binary variates and compute all pairwise MMDs by adding, say, 1, to each trait contribution, $(T_{1i} - T_{2i})^2 - 1/N_{1i} - 1/N_{2i} + 1 \approx (T_{1i} - T_{2i})^2 + 1$, since $1/N_{1i} \approx 0$ and $1/N_{2i} \approx 0$. Then samples are randomly drawn from the populations, the MMDs are estimated via Ossenberg's et al. (2006) suggestion, $(t_{1i} - t_{2i})^2 - 1/n_{1i} - 1/n_{2i} + 1$, and averaged. In this technique, all the MMD values are positive and, since the MMD is an unbiased estimator, population and averaged sample MMDs converge. It is evident that under Ossenberg's et al. (2006) correction all the MMDs are raised by the amount of r (when $const=1$) in relation to the population MMDs computed without correction. If we now alter the $const$ value to 2, 3, 4, …, we obtain a spectrum of MMDs, where for each $const$ value, population and averaged MMDs will converge, they will be positive numbers, and they are raised by the amount of r in relation to the MMDs of the previous $const$ value. Moreover, all MMDs at the various $const$ values give the same dendrograms, metric and non-metric multidimensional scaling (MDS) plots. Therefore, the choice of $const$ affects the magnitude of the MMD between two samples but it does not affect the differences among the pairwise MMD values of a dataset; it is these differences that give useful information about the relative biodistance among the populations and help to identify patterns and relationships between populations.
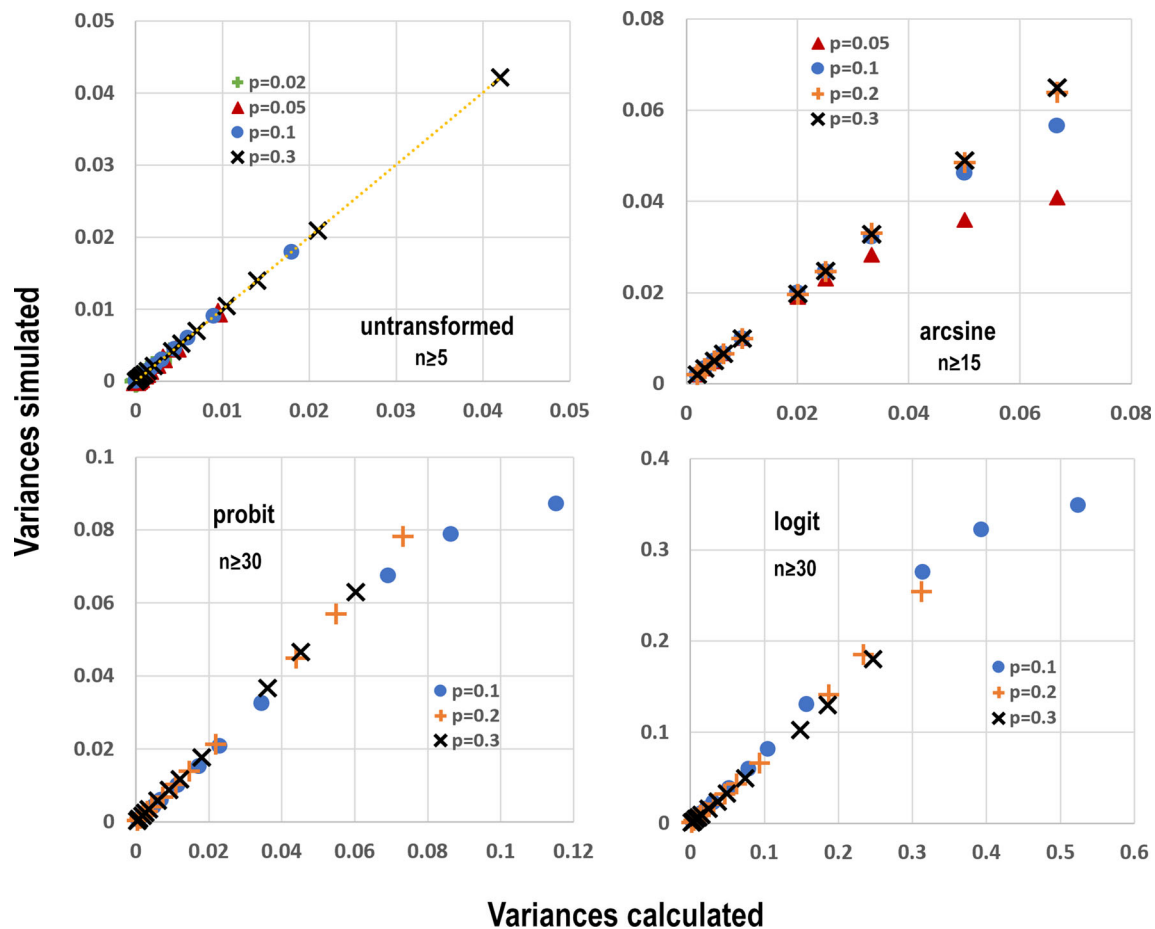
Fig. 1 Plots of simulated vs. calculated variances for untransformed data and data after arcsine, probit, and logit transformation

The situation is expected to be the same when *const* takes the values 0, −1, −2, … In this case, there is a *const* value that makes all MMDs negative. However, even in this case, population and averaged MMDs converge and they differ by a constant value from the corresponding positive MMDs computed using a high positive *const* value. Thus, the relative biodistances among the populations remain the same, and for this reason, the negative MMDs give precisely the same dendrogram as that of the positive MMDs. There is only a problem with the applicability of metric and nonmetric MDS because these techniques demand the input of a distance or dissimilarity matrix with non-negative elements. However, even in the case where all MMDs are negative, metric MDS can be applied using the cmdscale() R function, provided that the dimensions are selected based on the (two) smallest eigenvalues. Thus, the correction of negative MMDs concerns exclusively the applicability of nonmetric MDS and the solution proposed by Ossenberg et al. (2006) is the simplest one which solves the problem and it does not bias the distances.

As an example, consider Fig. 2, which shows MMDs computed on a simulated dataset using $r=20$, unit correlation matrix, step=0.05, $p1=0.2$, and $p2=0.4$. The sample sizes are 10, 20, 30, 40, and 50 and each population has a size of 10000. Note that since $p1=0.2$ and $p2=0.4$, the sample φ values and the corresponding population p values will range from 0.2 to 0.6 (0.4 plus 5 times 0.05, which is the step value). For this reason, the computed MMDs are positive. However, if we examine the individual trait contributions, we find that in almost all cases, there are negative contributions that affect the final MMDs shown in Fig. 2. In addition, due to these p values (which are greater than 0.1), the MMD is an unbiased estimator of population divergence despite the small sample sizes. This is shown in Fig. 2, where the values averaged over 500 MMDs of samples (o) randomly drawn from the populations converge satisfactorily to the corresponding population MMDs (•) when const = 0 (Fig. 2a) and const = 1 (Fig. 2b). Some small deviations concern only the very small first sample ($n = 10$). In contrast, if we transform the negative trait contributions to zeros, the population MMDs remain unaffected because there are no negative MMDs (•), whereas the
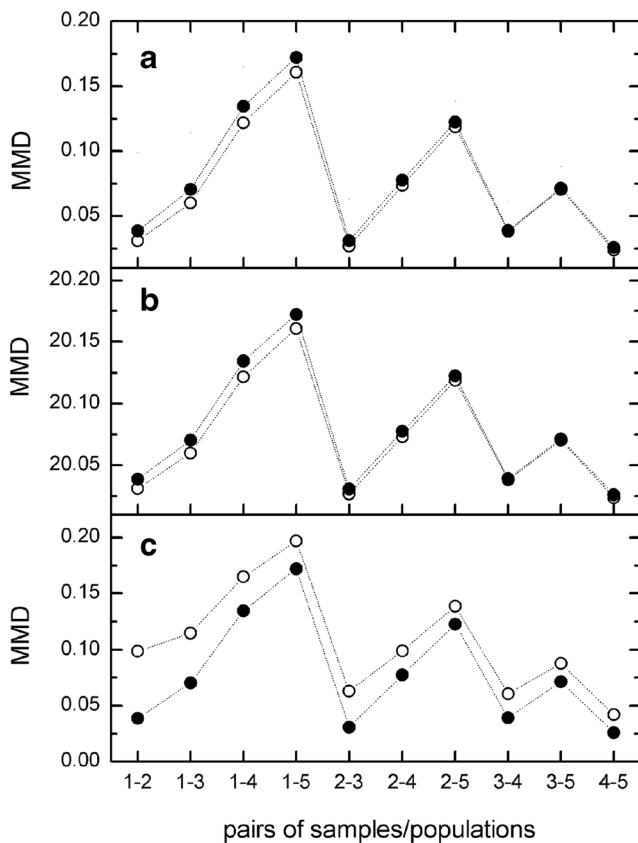
Fig. 2 Comparisons of MMDs when estimated on simulated populations (●) and averaging 500 distances of samples randomly drawn from these populations (o). Distances have been corrected for negative contributions as follows: No correction has been made in **a**; the value 1 has been added to all distances at each trait in **b**; and negative trait contributions have been transformed to zeros in **c**
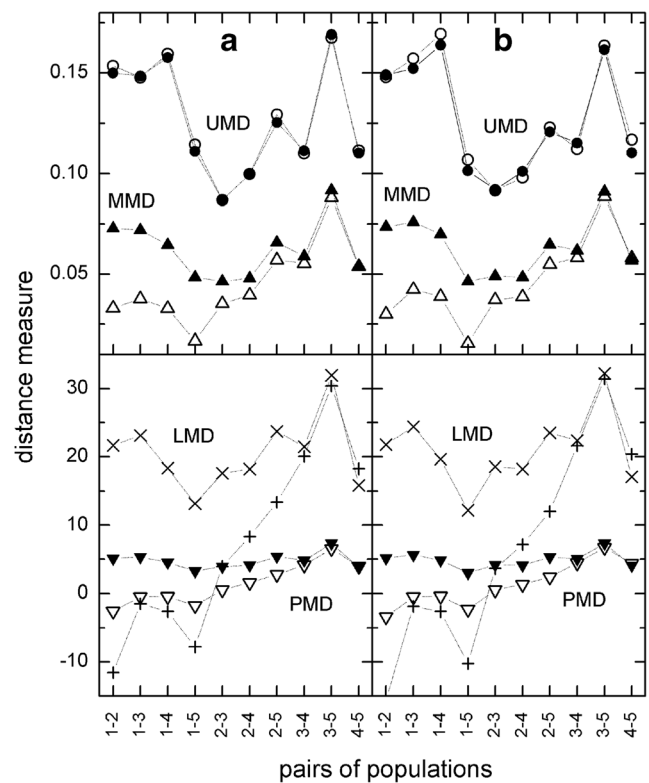

Fig. 3 Comparisons of the various distance measures when estimated on simulated populations, UMD (●), MMD (▲), LMD (×), PMD (▼), and averaging 1000 distances of samples randomly drawn from these populations UMD (o), MMD (△), LMD (+), PMD (▽). Binary datasets created using r=20, step=0.005, $p1$=0, $p2$=0.2, unit correlation matrix (**a**), and symmetric correlation matrix with the strongest positive correlations, (**b**). Population size=10000 and sample sizes: (1) 10, (2) 20, (3) 30, (4) 40, and (5) 50

corrected MMDs (o) are substantially raised in relation to the corresponding population MMDs (Fig. 2c), showing a clear bias of the computed MMD values.

The discussion presented above was focused on the MMD using the arcsine transformation. However, it is evident that it concerns all the measures of divergence (MDs) presented in this study (UMD, PMD, LMD). Moreover, since negative MMDs concern exclusively the applicability of metric and nonmetric MDS and the present study does not involve multidimensional scaling techniques, no correction for negative MDs has been applied in the remainder of the paper.

Results from simulated data

Representative results obtained from the simulated data are presented in the Supporting Data S2. This file includes tables and figures that show the effect of sample size and the presence of traits with very small φ values on whether the distances under study are unbiased estimators of population divergence or not. The simulated data was created using function *simbiasbinMDs*, which also performed the analysis and

created dendrograms. Figures 3 and 4 present comparisons of the distance measures UMD, MMD, PMD, and LMD when the binary datasets have been created using unit correlation matrix (i.e., a matrix with almost zero correlations) and a correlation matrix with the strongest positive correlations. Note that the histograms of the Pearson correlation coefficients that correspond to the correlation matrices of Figs. 3 and 4, given in the Supporting Data S2, show a peak at 0.75. In Fig. 3, the sample sizes are small (10-20-30-40-50), whereas in Fig. 4, the sample sizes range from 100 to 500 (100-200-300-400-500) cases. In both figures, the percentage of traits with low relative frequencies is 41% for $p < 0.1$ and 16% for $p < 0.05$.

It is seen that both sample size and the presence of traits with very small φ values play an important role in whether a distance is an unbiased estimator of population divergence or not. As a rule, when the sample sizes are relatively large, greater than 100 or 200, even for percentages of traits as large as 90% for $p < 0.1$ and 40% for $p < 0.05$, all the examined distances are either unbiased or nearly unbiased estimators of population divergence. The presence of strong intercorrelations practically does not affect this property. As the sample
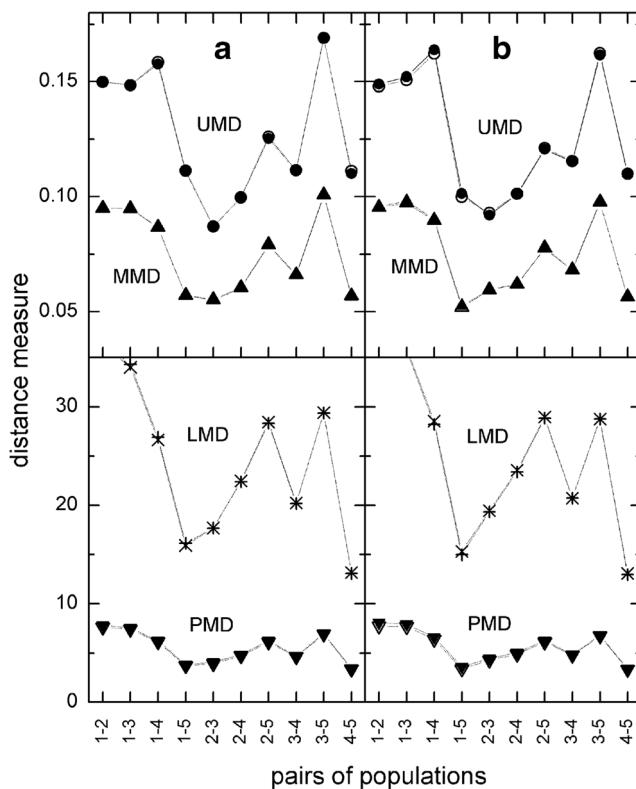
Fig. 4 Comparisons of the various distance measures when estimated on simulated populations, UMD (●), MMD (▲), LMD (×), PMD (▼), and averaging 1000 distances of samples randomly drawn from these populations UMD (○), MMD (△), LMD (+), PMD (▽). Binary datasets created using $r=20$, step=0.005, $p1=0$, $p2=0.2$, unit correlation matrix (**a**), and symmetric correlation matrix with the strongest positive correlations, (**b**). Population size=10000 and sample sizes: (1) 100, (2) 200, (3) 300, (4) 400, and (5) 500

sizes decrease, below 50 cases per sample, all distances, except the UMD, start to show a biased estimation of the population divergence. Note that some, usually small, deviations between population and averaged UMDs disappear when we increase the repetition number. Therefore, the UMD appears to be the most robust unbiased estimator of population divergence among all measures of divergence studied. Again, the intercorrelations among the traits do not affect the biased/unbiased estimation of the population divergence. This is an important result because it allows us to simulate a specific dataset of multivariate binary data using the simple unit correlation matrix.

Results from dental nonmetric traits of archeological assemblages

The results obtained from the dental nonmetric traits are shown in Figs. 5 and 6 and the Supporting Data S3. The calculations in this spreadsheet were performed using the *biasbinMDs* function. As mentioned above, this function assesses biases in the estimation of population divergence by using the original dataset to simulate the respective

populations and subsequently calculating pairwise population distances and pairwise distances between samples randomly drawn from the populations. The comparison between the population and averaged samples pairwise distances allows the assessment of which of the distances are unbiased estimators of population divergence.

The original GR-28 dataset includes many traits with very low frequencies (Table 1). However, despite the presence of a relatively large proportion of low frequencies (21%), the UMD is an unbiased estimator (Fig. 5). The MMD, PMD, and LMD exhibit biased estimations of the population divergence, although the bias in the MMD is smaller than that in the PMD and LMD (Fig. 5 and Supporting Data S3). In addition, we observe that most of the values of the PMD and LMD are negative but, as explained above, this is not a problem and it can be addressed via Ossenberg's et al. (2006) correction.

If we use data editing and remove traits with $p < 0.01$, we obtain the GR-12 dataset, i.e., a dataset with 16 traits less. This is a great trait reduction and is likely to lead to a significant loss of information about the clustering of samples/populations. In the reduced dataset, both the UMD and MMD are unbiased estimators of population divergence and this property has been considerably improved for the other distances as well (Fig. 5 and Supporting Data S3).

The potential impact of the loss of information when reducing the dataset can be seen using dendrograms. Figure 6 shows Ward's dendrograms obtained from the UMD and MMD when applied to the GR-28 and GR-12 datasets. It is seen that there is indeed a loss of information about clusters when passing from GR-28 to GR-12. For UMD and MMD, the dendrograms of the GR-28 dataset show two broad clusters, one encompassing all prehistoric assemblages (Petras, Livari, Xeropigado, Pella) and the other including the historical materials (Christianoi, Eleutherna, Akraiphnio). When using GR-12, the picture obtained is very similar but now the Early Bronze Age assemblage from Pella clusters together with the historical materials, which is very difficult to explain archeologically.

These very tentative bioarcheological results show that MMD and UMD provide largely the same information but UMD outperforms the other measures as an unbiased estimator of population divergence, as expected based on the mathematical principles underlying its definition. With regard to whether this measure also generates meaningful biodistances, in our dataset, this is the case; however, further research employing larger and more diverse archeological assemblages which form known biodistance clusters is needed.

Results concerning *p* value estimation

The results concerning the *p* values are given in Fig. 7 and in the Supporting Data S2 and Supporting Data S3 and they have been obtained using the *simperbinMDs* and *perbinMDs*

Fig. 5 Comparisons of UMD and MMD when estimated **a** on GR-28 and GR-12 datasets of binary data (+), **b** on the corresponding simulated populations (o), and **c** by averaging the distances of 1000 samples randomly drawn from the populations (×). Population and averaged distances are corrected by subtracting the corresponding variances
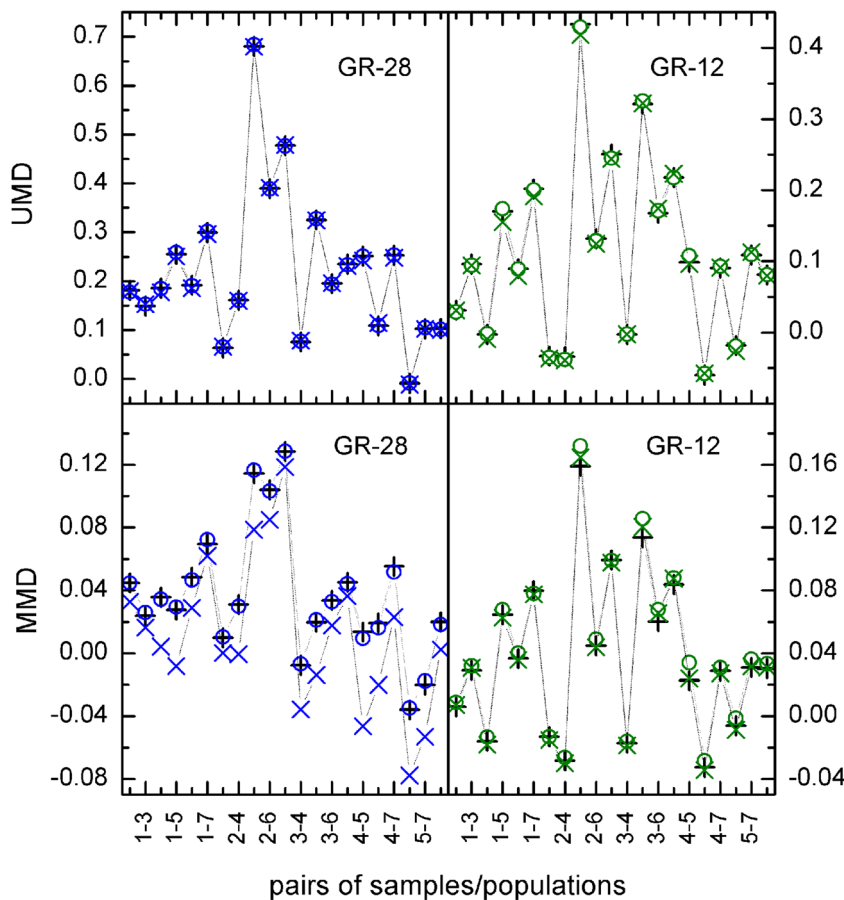
functions. The comparisons between *p* values calculated from the test statistics S and T and the permutation method reveal the following. There is a characteristic difference between the MMD, UMD, and the rest of the measures of divergence concerning the pattern of their *p* values. Specifically, when the PMD and LMD take negative values, whereas the corresponding MMD, UMD are positive, the *p* values calculated from the test statistics S and T are much greater that those

estimated from the permutation method, which gives *p* values similar to those corresponding to the MMD and UMD. This is an indication that under these conditions, the permutation method is more reliable than the test statistics.

In what concerns the MMD and UMD, although the pattern of their *p* values is overall the same, there are differences in the *p* values calculated from the various techniques, i.e., from the test statistics S and T, Equations (8) and (9), and the Monte-
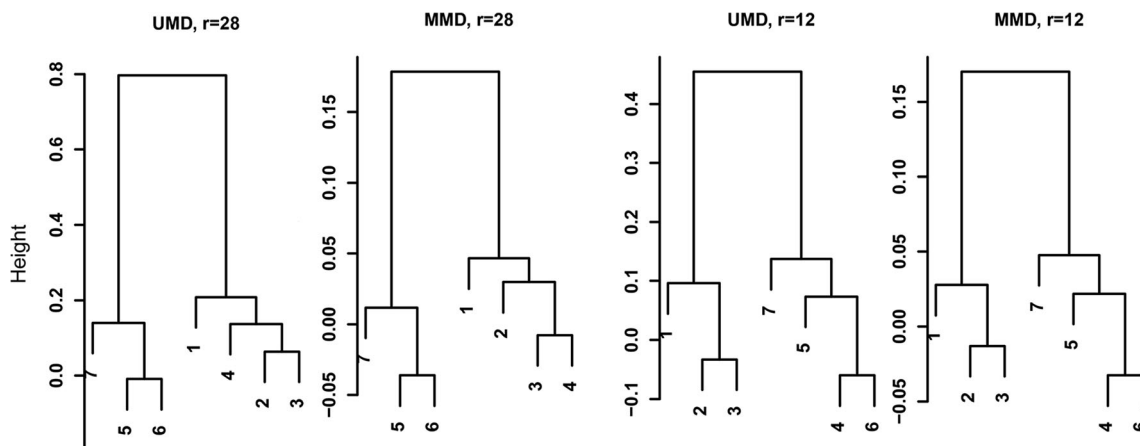

Fig. 6 Ward's dendrograms obtained from the UMD and MMD when applied to the GR-28 and GR-12 datasets. Key: 1 = Petras, 2 = Livari, 3 = Xeropigado, 4 = Pella, 5 = Christianoi, 6 = Eleutherna, 7 = Akraiphnio
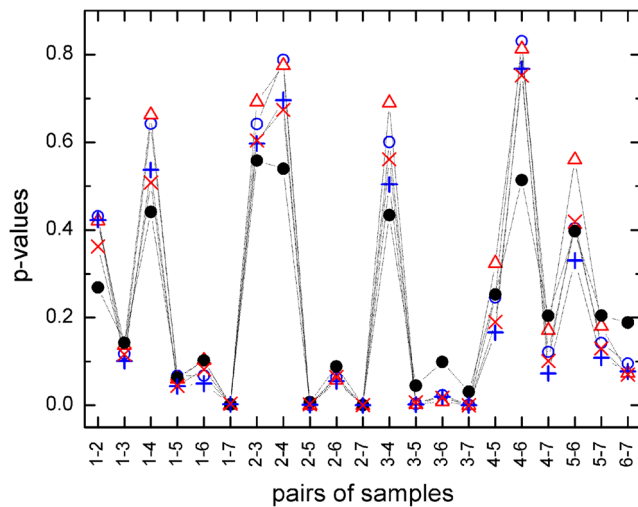
Fig. 7 *p* values for the MMD of the GR-12 dataset estimated from test statistics S (o) and T (△) using Equations (8) and (9) and the permutation method using as test statistic the values *d* of the MMD (•) as well as the test statistics S (+) and T (×) of Equations (8) and (9), respectively

Carlo permutation method based on the S, T, and d statistics (Fig. 7). Nevertheless, we have not detected significant differences between *p* values calculated from the test statistics S, T and the permutation method, even in the analysis of simulated data with strong intercorrelations. Note that the test statistics S and T are valid if the transformed data is nearly normally distributed and all *r* traits are independent. Thus, the comparisons of the *p* values do not show which of the techniques examined provide the best estimation of the *p* value. However, at this point, we should clarify that the calculation of different *p* values for a certain distance when all *p* values are say above 0.1 is unimportant; on the contrary, it shows that there is no evidence that the distance is statistically significant. Similarly, different *p* values but all well below 0.05 provide strong evidence against the null hypothesis and, therefore, indicate statistical significance since the smaller the *p* value, the stronger the evidence for statistical significance. Finally, in the limiting case where the estimated *p* values lie around 0.05, it is up to the researcher to decide about distance significance. From this point of view, in most cases concerning the MMD and UMD, the differences among the *p* values calculated from the various techniques examined in this paper are not so substantial to prevent us from reaching safe conclusions about the significance of distances under study. In contrast, the estimation of the significance of the computed distances based on all techniques provides a more secure result. For the remaining distances, the significance should be based mainly on the permutation method.

A revision of the data editing procedure

Currently, the MMD is usually applied after data editing; naturally, the same could be done for all distances presented

above, especially the PMD and LMD. The data editing procedure involves the elimination of the following traits from the dataset:

1. Traits that exhibit only missing values in one or more samples under study.
2. Traits that exhibit a particularly high (>0.95) or low (<0.05) frequency within one or more samples under study.
3. Nondiagnostic traits, that is, traits that are not significantly different between at least one pair of samples (Harris and Sjøvold 2004).
4. Traits that exhibit a statistically significant intercorrelation with one another (Irish 2010 and references therein).

Based on the present study, this procedure should be revised except for the first step since none of the measures of divergence described above can be computed if there are traits that exhibit only missing values in one or more samples. For the second step, traits that exhibit a particularly high or low frequency within one or more samples may cause problems since, apart from the untransformed data, all the other expressions for the variance of the transformed data are approximately valid within a certain range of *p* and *n* values. Thus, the arcsine transformation fails for $p \leq 0.1$ or $p \geq 0.9$ when $n \leq 20$, whereas for probit/logit, these limits become $p \leq 0.15$ or $p \geq 0.85$ when $n < 30$. However, the present study showed that these limits may be violated and a measure of divergence can show small biased estimations of the population divergence. This is especially true for the MMD. In any case, the second step may be revised as follows. We examine whether the distance under examination is a biased or an unbiased estimator of population divergence. If it is an unbiased estimator, there is no need to delete any traits. If it is a biased estimator, we remove trait(s) starting from those with the smallest/highest frequency until we obtain a dataset in which the computed distance is an unbiased or nearly unbiased estimator of population divergence. This process may be easily implemented using function *biasbinMDs*.

Concerning the third step listed above, it may be ignored. Nondiagnostic traits may favor the appearance of negative MDs. However, the existence of such traits indicates that the populations under study are biologically close to each other, whereas their elimination enhances the differences between populations, which may yield false conclusions about their biodistance. If a researcher wants to avoid negative distance values, Ossenberg et al. (2006) offer the best solution. Finally, in what concerns the presence of strongly intercorrelated traits, as we found using simulated data, such traits do not yield biased estimations of population divergence and do not have a significant effect on the *p* values estimated from the S statistic, Equation (8), which assumes that all *r* traits should be independent. Note that in a 2015 study, it was also shown that the inclusion of intercorrelated traits does not appear to affect the

validity of the MMD results (Nikita 2015). However, the main reason for many researchers to delete intercorrelated traits is that the MMD does not take into account the effect of intercorrelated traits on the population divergence. From this point of view, this is a justifiable procedure, which, though, as any data elimination approach, may lead to biased MMD values.

Conclusions

In the present study, we examined three new measures of divergence based on untransformed data (UMD) and the logit (LMD) and probit (PMD) transformations. In addition, we examined the conventional Smith's mean measure of divergence (MMD) based on the arcsine transformation. The main conclusions that can be drawn are the following:

1. The UMD based on untransformed data outperforms the other measures. It is an unbiased estimator of population divergence and does not exhibit application problems at very low or very high frequencies.
2. The MMD is a satisfactory distance measure for binary data, although its application requires a careful test to avoid biased estimations of population divergence when there are small sample sizes and many traits with frequencies lower than 0.1 or/and greater than 0.9.
3. The UMD and MMD usually give similar information about the relative biodistance and the existing patterns and relationships between populations.
4. The measures of divergence based on the probit and logit transformations are more prone to biased estimations of population divergence than the UMD and MMD, especially in datasets with small sample sizes and traits with very low/high frequencies. Since they have no advantage over the UMD and MMD, these measures may be considered inappropriate for the study of datasets of dental (as well as cranial) traits.
5. The statistical significance of the estimated distances should be based on both the S and T test statistics and the permutation method avoiding assumptions on trait independence and normality of transformed/untransformed data.
6. The conventional data editing procedure should be revised. It should be mainly related to the need for a measure of divergence to be an unbiased estimator. If a measure of divergence is an unbiased estimator for a certain dataset, there is no need to delete any traits. If though this measure is a biased estimator, we remove trait(s) starting from those with the smallest/highest frequency until we obtain a dataset in which the computed measure becomes an unbiased estimator of population divergence.

Declarations

Competing interests The authors declare no competing interests.

References

Akamatis I (2009) Prehistoric Pella: Bronze Age cemetery. In: Drougou S, Evgenidou D, Kritzas C, Kaltsas N, Penna V, Tsourti I, Galani-Krikou M, Ralli E (eds) Kermatia Filias: Honorary Volume for Ioannis Touratsoglou. Numismatic Museum, Athens, pp 193–213 (in Greek)

Anscombe FJ (1948) The transformation of Poisson, binomial and negative-binomial data. Biometrika 35:246–254. https://doi.org/10.2307/2332343

Bartlett MS (1936) The square root transformation in the analysis of variance. Suppl J R Stat Soc 3:68–78. https://doi.org/10.2307/2983678

Bartlett MS (1947) The use of transformations. Biometrics 3:39–52. https://doi.org/10.2307/3001536

Berry AC, Berry RJ (1967) Epigenetic variation in the human cranium. J Anat 101:361–379

Berry AC (1974) The use of non-metrical variations of the cranium in the study of Scandinavian population movements. Am J Phys Anthropol 40:345–358. https://doi.org/10.1002/ajpa.1330400306

Berry AC (1976) The anthropological value of minor variants of the dental crown. Am J Phys Anthropol 45:257–268. https://doi.org/10.1002/ajpa.1330450211

Bertsatos A, Chovalopoulou ME (2016) A GNU Octave function for Smith's mean measure of divergence. Bioarchaeol Near East 10:69–73

Blumenbach JF (1775) De Generis Humani Varietate Nativa. Friedrich Schiller University, Jena

Bourbou C (2004) The people of Early Byzantine Eleutherna and Messene (6th–7th centuries AD). A bioarchaeological approach. Dissertation, University of Crete. (in Greek)

Broca MP (1863) Review of the proceedings of the anthropological society of Paris. Anthropol Rev 1:274–310

Buikstra JE (1979) Contributions of physical anthropologists to the concept of Hopewell: a historical perspective. In: Brose BS, Greber N (eds) Hopewell Archaeology. Kent State University Press, Kent, pp 220–233

Buikstra JE, Frankenberg SR, Konigsberg LW (1990) Skeletal biological distance studies in American physical anthropology. Am J Phys Anthropol 82:1–7. https://doi.org/10.1002/ajpa.1330820102

Corruccini RS (1972) The biological relationship of some prehistoric and historic Pueblo populations. Am J Phys Anthropol 37:373–388. https://doi.org/10.1002/ajpa.1330370307

Freeman MF, Tukey JW (1950) Transformations related to the angular and square root. Ann Math Stat 21:607–611

Green RF, Suchey JM (1976) The use of inverse sine transformations in the analysis of non-metric cranial data. Am J Phys Anthropol 45:61–68. https://doi.org/10.1002/ajpa.1330450108

Grewal MS (1962) The rate of genetic divergence in the C57BL strain of mice. Genet Res 3:226–237. https://doi.org/10.1017/S0016672300035011

Harris EF, Sjøvold T (2004) Calculation of Smith's mean measure of divergence for inter-group comparisons using nonmetric data. Dental Anthropol 17:83–93

Hefner JT, Pilloud MA, Buikstra JE, Vogelsberg CCM (2016) A brief history of biological distance analysis. In: Pilloud MA, Hefner JT (eds) Biological distance analysis: forensic and bioarchaeological perspectives. Academic Press, San Diego, pp 1–22

Irish JD (2010) The mean measure of divergence: its utility in model-free and model-bound analyses relative to the Mahalanobis D2 distance for nonmetric traits. Am J Hum Biol 22:378–395. https://doi.org/10.1002/ajhb.21010

Irish JD (2016) Who were they really? Model-free and model-bound dental nonmetric analyses to affirm documented population affiliations of seven South African "Bantu" samples. Am J Phys Anthropol 159:655–670. https://doi.org/10.1002/ajpa.22928

Kalliga E (2015) Prehistoric cemeteries of Macedonia. The Early Bronze Age cemetery at Pella: the analysis of human skeletal remains. Dissertation, Aristotle University of Thessaloniki. (in Greek)

Kappas M, Sakkari S (2012) Christianoi. Archaiologikon Deltion B1, Chronika (in Greek)

Maniatis Y, Ziota C (2011) Systematic ^{14}C dating of a unique Early and Middle Bronze Age cemetery at Xeropigado Koiladas, West Macedonia, Greece. Radiocarbon 53:461–478. https://doi.org/10.1017/S0033822200034597

Nikita E, Mattingly D, Lahr MM (2012) Sahara: barrier or corridor? Nonmetric cranial traits and biological affinities of North African Late Holocene populations. Am J Phys Anthropol 147:280–292. https://doi.org/10.1002/ajpa.21645

Nikita E (2015) A critical review of the mean measure of divergence and Mahalanobis distances using artificial data and new approaches to the estimation of biodistances employing nonmetric traits. Am J Phys Anthropol 157:284–294. https://doi.org/10.1002/ajpa.22708

Nikita E (2017) Osteoarchaeology: A guide to the macroscopic study of human skeletal remains. Academic Press, San Diego

Nikita E, Schrock C, Sabetai V, Vlachogianni E (2019) Bioarchaeological perspectives to diachronic life quality and mobility in ancient Boeotia, central Greece: preliminary insights from Akraiphia. Int J Osteoarchaeol 29:26–35. https://doi.org/10.1002/oa.2710

Ossenberg NS, Dodo Y, Maeda T, Kawakubo Y (2006) Ethnogenesis and craniofacial change in Japan from the perspective of nonmetric traits. Anthropol Sci 114:99–115. https://doi.org/10.1537/ase.00090

Papadatos Y, Sofianou Ch (2015). Livari Skiadi. A Minoan cemetery in Lefki, Southeast Crete. Volume I: excavation and finds. Prehistory Monographs 50, Philadelphia.

Pietrusewsky M (2013) Biological distance in bioarchaeology and human osteology. In: Smith C (ed) Encyclopedia of global archaeology. Springer, Berlin, pp 889–902

Pilloud MA, Hefner JT (2016) Biological distance analysis. Academic Press, San Diego, Forensic and bioarchaeological perspectives

Relethford JH (2016) Biological distances and population genetics in bioarchaeology. In: Pilloud MA, Hefner JT (eds) Biological distance analysis: forensic and bioarchaeological perspectives. Academic Press, San Diego, pp 23–34

Rightmire GP (1970) Bushman, Hottentot and South African Negro crania studied by distances and discrimination. Am J Phys Anthropol 33:169–196. https://doi.org/10.1002/ajpa.1330330204

Sabetai V (1995) Akraiphia. Archaiologikon Deltion B1. Chronika 50: 301–304 (in Greek)

Santos F (2018) AnthropMMD: An R package with a graphical user interface for the mean measure of divergence. Am J Phys Anthropol 165:200–205. https://doi.org/10.1002/ajpa.23336

Sjøvold T (1973) The occurrence of minor non-metrical variants in the skeleton and their quantitative treatment for population comparisons. Homo 24:204–233

Sjøvold T (1977) Non-metrical divergence between skeletal populations. Ossa 4(suppl):1–133

Sołtysiak A (2011) An R script for Smith's mean measure of divergence. Bioarchaeol Near East 5:41–44

Souza P, Houghton P (1977) The mean measure of divergence and the use of non-metric data in the estimation of biological distances. J Archaeol Sci 4:163–169. https://doi.org/10.1016/0305-4403(77)90063-2

Stojanowski CM (2019) Ancient migrations: Biodistance, genetics, and the persistence of typological thinking. In: Buikstra JE (ed) Bioarchaeologists speak out. Deep time perspectives on contemporary issues. Springer, Cham, pp 181–200

Themelis P (1994–1996) Eleutherna. Kretiki Estia 5:270–283 (in Greek)

Triantaphyllou S (2010) Aspects of life histories from the Bronze Age cemetery at Xeropigado Koiladas, western Macedonia. In: Philippa-Touchais A, Touchais G, Voutsaki S, Wright J (eds) Mesohelladika: La Grèce continentale au Bronze Moyen. Bulletin de Correspondance Helléniques Suppl. 52, Athènes, pp 969–974

Triantaphyllou S (2012) Kephala Petras Siteias: human bones and burial practices in the EM rockshelter. In: Tsipopoulou M (ed) Petras-Siteia, 25 years of excavations and studies, conference organized by the Danish Institute, Athens, 9-10 October 2010. Aarhus University Press, Århus, pp 161–166

Triantaphyllou S (2016) Staging the manipulation of the dead in Pre-and Protopalatial Crete, Greece (3rd–early 2nd mill. BCE): from body wholes to fragmented body parts. J Archaeol Sci Rep 10:769–779. https://doi.org/10.1016/j.jasrep.2016.06.003

Tsipopoulou M (2010) Prepalatial rockshelter at Petras Siteias – first announcement. In: Andrianakis M, Tzachili I (eds) The Archaeological Work in Crete 1, Proceedings of the 1st Meeting, Rethymnon 28-30 November 2008. Faculty of Letters Publications, University of Crete, Rethymnon, pp 121–138 (in Greek)

Ullinger JM, Sheridan SG, Hawkey DE, Turner CG, Cooley R (2005) Bioarchaeological analysis of cultural transition in the southern Levant using dental nonmetric traits. Am J Phys Anthropol 128: 466–476. https://doi.org/10.1002/ajpa.20074

Vlachogianni E (1997) Akraiphia. Archaiologikon Deltion B1. Chronika 52:377–392 (in Greek)

Zertuche F, Meza-Peñaloza A (2020) A parametric bootstrap for the Mean Measure of Divergence. Int J Biostat 16(2). https://doi.org/10.1515/ijb-2019-0117