Dentitions, Distance, and Difficulty: A Comparison of Two Statistical Techniques for Dental Morphological Data

Heather Joy Hecht Edgar*

Laboratory of Human Osteology, Maxwell Museum of Anthropology, University of New Mexico, Albuquerque, New Mexioc 87131

ABSTRACT   One of the main uses of dental morphological data is to study patterns of affinities among populations. Many different approaches to this purpose are available, each one having its own strengths and weaknesses. For this study, observations were made of the morphology of 614 African American and 327 European American dentitions (n = 941). Each of these samples was divided into three groups based on the time in which they lived. Affinities among the resulting six groups were estimated based on the frequencies of dental morphological characteristics, by the use of both the Mean Measure of Divergence and a Pseudo-Mahalanobis' D^2. The results of these analyses are compared using a Procrustes transformation that rotates and scales coordinates derived from distances until achieving the best fit. The two statistics produce similar, although not identical results. The appropriate use and relative value of each approach is discussed. *Dental Anthropology* 2004;17(2):55-62.

The Mean Measure of Divergence (MMD) has become the standard statistical technique for assessing biological affinities when using frequencies of dental morphological characteristics (Scott and Turner, 1997). There are several advantages in using this statistic: It is appropriate for nominal data, it is relatively easy to compute, and it is comparable among researchers. There is however, a drawback to using the MMD; it is only appropriately used when the traits being studied are independent. The assumption of independence is weak for several dental characteristics, so inter-trait correlations must be tested, and traits that are correlated must be removed from of a MMD analysis.

An alternative to MMD is the Mahalanobis' D^2 statistic, which allows correlated features to be used in affinity measures (Mahalanobis, 1936). However, as originally formulated, this statistic is useful only for metric, not nominal, data. Konigsberg (1990) used a pseudo-Mahalanobis' D^2 to determine biological affinity using non-metric data. This statistic has the potential to allow distance measures to be based on a greater variety and number of dental characteristics than the MMD. Of course, like MMD, the D^2 statistic has its drawbacks. The primary problems with the application of this statistic are its limited applicability when analyzing a number of traits with little or no correlation, the need for multiple observations per individual, and its relatively more difficult computation. Because every trait must be compared to every other for each sample being studied, comparing more than a few traits at a time can become quite arduous, even with a computer. Additionally, the inclusion of a new sample for analysis requires the recalculation of all measures of affinity among groups, not simply the measures of affinity of the new sample with the original groups, as with the MMD.

This study presents the results of a comparison of MMD and pseudo-D^2 methods for determining biological affinity among several samples. The goals are to investigate whether the two types of analysis result in similar findings, and if not, to consider why.

MATERIAL

The data for this study comes from the dentitions of 941 African Americans and European Americans, analyzed as part of a larger study of the microevolution of African American dental morphology. Samples come from collections temporarily or permanently housed at the National Museum of Natural History, National Museum of Health and Medicine, Cleveland Museum of Natural History, University of Tennessee Health Sciences Center, Ohio State University, and Arizona State University. The samples were divided into six groups, based on ancestry and time period. The samples sizes and time periods are listed in Table

Address for correspondence: Heather J. H. Edgar, Laboratory of Human Osteology, Maxwell Museum of Anthropology, University of New Mexico, Albuquerque, NM 87131.
Email: hjhedgar@unm.edu

Table 1. Sample compositions

	Early born circa 1650-1850	Middle born circa 1825-1910	Late born circa 1920-1960	total
African American	35	414	165	614
European American	33	139	155	327
total	68	553	320	941

1.

For this study, a maximum of 136 observations of 32 morphological characteristics was possible per dentition. Observation procedures were based on the Arizona State University dental anthropology system (Turner *et al.*, 1991). No significant directional asymmetry of expression or sexual dimorphism was found, so rights and lefts were combined (with the greatest trait expression being represented), as were observations from males and females. Observations were then dichotomized with guidance from Haeussler *et al.* (1989), Irish (1993), Irish and Turner (1990), Scott and Turner (1997), and Turner (1987).

All statistics were performed using the SAS statistical package (SAS Institute Inc., 1990). Associations between traits were determined using the likelihood ratio statistic. The list of traits that was used for each analysis can be found in Table 2. Traits used in the MMD analysis are independent from each other. To invert the matrix of correlations, the D^2 analysis requires that most variables have some

tetrachoric correlation with all other variables. Several variables were eliminated from D^2 analyses because they were found to have little or no correlation with other variables, and thus the tetrachoric correlation matrix was singular. Different variable combinations were used in each analysis because of the requirements of each statistics; traits should be uncorrelated for the MMD and correlated for the D^2.

STATISTICAL METHODS

Mean Measure of Divergence

The MMD statistic was developed by C. A. B. Smith, and was first used to look at changes due to inbreeding in mice (Grewal, 1962; Berry *et al.*, 1967). Berry and Berry (1967) first applied it to the study of biological affinities or distance in humans. The MMD estimates biological distance between samples based on the degree of phenetic similarity (Irish, 1997). The statistic requires an assumption of independence of traits. Like D^2, it is useful if trait expression varies in a population, when frequencies are 5-95% (de Souza and Houghton, 1977). Some major benefits of its use are its ability to work with incomplete data and its applicability to samples as small as 10-20 observations. MMD is defined as:

$$\text{MMD} = (\sum(\Theta_1 - \Theta_2)^2 - (1/n_1 + 1/n_2))/c$$

where Θ_1 and Θ_2 are the arc sin ($\sin^{-1}$) transformations of the observed frequencies in the two samples being compared, n_1 and n_2 are the sample sizes, and c is the number of characters employed (Freeman and Tukey, 1950).

Pseudo-Mahalanobis' D^2

The Pseudo-Mahalanobis' D^2 is defined as the sum of squares of differences between corresponding mean values of two sets of measurements, weighted by the variance/covariance matrix (Burnaby, 1966):

$$D^2 = (\chi_{ik} - \chi_{jk})' \sum (\chi_{ik} - \chi_{jk})$$

where χ_{ik} is the mean of expression for sample i for k traits, and χ_{jk} is the same for sample j. The middle term ($\sum$) is the pooled covariance matrix between the k traits (Manly, 1994). In this study, the means of trait expressions are the threshold values corresponding to the trait frequencies in the samples (Falconer, 1981), and the middle term is a pooled matrix of tetrachoric correlations between the traits (Brown, 1977). These transformations account for correlations between characteristics being used (Konigsburg, 1990; Mizoguchi, 1977) and the threshold nature of dental morphological traits (Scott and Turner, 1997).

Table 2. Dental characters used in each analysis

Max MMD	Mand MMD	Max D²	Mand D²
DIAS	LI2SS	UI2SS	LI1SS
UCSS	LCDR	UCSS	LI2SS
UI1LC	LP3LC	UI1LC	LP4LC
UI2DS	LP4LC	UI2TD	LM2MT
UI2IG	LM1AF	UCTD	LM1PS
UM3CA	LM2GP	UCDR	LM2PS
UCTD	LM1DW	UP3MD	LM2C5
UCDR	LM1MT	UP4MD	LM1C6
UP3MD	LM2PS	UM1MC	
UP4MD	LM2C5	UM2MC	
UM2MC	LM1C6	UM1HC	
UM1HC	LM2C7	UM1C5	
UM2C5		UM2C5	
UM2CB		UM1CB	
		UM2CB	

Table 3. MMD distances, maxillary traits

	Late AA	Late EA	Middle AA	Middle EA	Early AA	Early EA
Late AA	0	0.113	0.074	0.443	0.244	0.402
Late EA	0.113	0	0.113	0.231	0.395	0.239
Middle AA	0.074	0.113	0	0.222	0.187	0.247
Middle EA	0.443	0.231	0.222	0	0.292	0
Early AA	0.244	0.395	0.187	0.292	0	0.218
Early EA	0.402	0.239	0.247	0.000	0.218	0

Procrustes' transformation

The purpose of this statistic is to rotate and scale two sets of coordinates so as to achieve the best fit between them (Gower, 1971, 1975). For this study, the coordinates come from principal coordinates analysis of four distance matrices, and represent the first two axes of each matrix. The better the fit between two sets of coordinates, the smaller the summed deviations should be. Gower (1971) refers to the statistic as R^2 (for residual), but it can also be found as S^2 (for sum of squares) (Goodall, 1991) and M^2 (for minimum)(Jackson, 1995). R^2 is defined as:

$$R^2 = \sum \Delta^2(P_i P_i^*),$$

where P_i and P_i^* represent the corresponding points in two different sets of coordinates. The R^2 statistic is the sum of squared differences after rotation and scaling. The smaller the R^2, the smaller the difference is between the two sets of coordinates. For this study, a small R^2 will indicate good agreement between the MMD and D^2 statistics.

RESULTS

Before discussing the direct comparison of statistical methods, an examination of the pictures presented by each analysis is in order. Due to the difficulty in performing pseudo-Mahalanobis' D^2 with a large quantity of traits, maxillary and mandibular traits were considered separately.

Measures of affinity

Results for MMD analyses based on maxillary and mandibular traits can be seen in Tables 3 and 4, respectively. The maxillary traits show a separation between African Americans (AA) and European Americans (EA) at all time periods. There is a closer relationship between early and middle EA than either to late EA. Early AA is different from all groups, with middle and late AA being most like late EA. Analysis of the mandibular traits emphasizes the split between EA and AA and minimizes other details.

Results for the D^2 analyses are summarized in Tables 5 (maxillary traits) and 6 (mandibular traits). The results for the maxillary traits seem to emphasize the time difference between groups rather than differences in ancestry. Late and middle AA and EA cluster most closely, with early AA and EA being very distant from each other and all other groups. The results based on the mandibular trait D^2 are the most difficult to characterize. There is a large difference between early and middle AA, and a relatively small difference between middle and late AA. While the indication that change in the African American gene pool slowed down after the Civil War reflects known historical patterns of admixture (Davis, 1991), it does not explain the apparent similarity of early EA and middle AA, the smallest distance in the matrix. This information is graphically presented in Figure 1, which shows the principal coordinates of the relationships among the six groups resulting from MMD analyses,

Table 4. MMD distances, mandibular traits

	Late AA	Late EA	Middle AA	Middle EA	Early AA	Early EA
Late AA	0	0.507	0.094	0.471	0.122	0.488
Late EA	0.507	0	0.525	0.119	0.601	0.148
Middle AA	0.094	0.525	0	0.401	0.122	0.374
Middle EA	0.471	0.119	0.401	0	0.449	0.000
Early AA	0.122	0.601	0.122	0.449	0	0.410
Early EA	0.488	0.148	0.374	0	0.410	0

Table 5. D² distances, maxillary traits

	Late AA	Late EA	Middle AA	Middle EA	Early AA	Early EA
Late AA	0	4.175	7.692	7.755	6.676	17.243
Late EA	4.175	0	4.472	4.563	10.015	10.769
Middle AA	7.692	4.472	0	3.184	7.982	8.698
Middle EA	7.755	4.563	3.184	0	8.303	6.499
Early AA	6.676	10.015	7.982	8.303	0	10.295
Early EA	17.243	10.763	8.698	6.499	10.295	0

Table 6. D² distances, mandibular traits

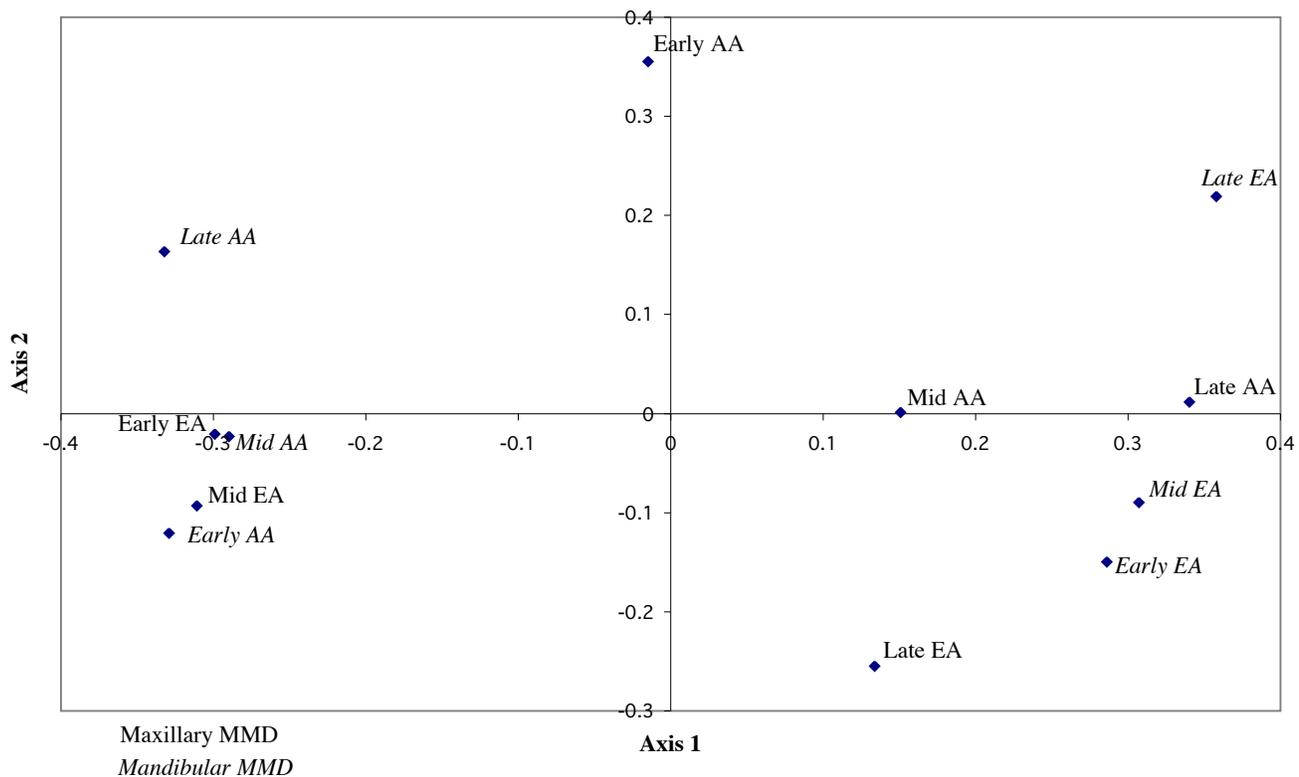|	Late AA	Late EA	Middle AA	Middle EA	Early AA	Early EA
Late AA	0	1.473	8.630	3.593	8.725	8.302
Late EA	1.473	0	4.598	4.714	6.300	5.243
Middle AA	8.630	4.598	0	8.442	7.281	2.442
Middle EA	3.593	4.714	8.442	0	5.040	7.459
Early AA	8.725	6.300	7.281	5.040	0	8.800
Early EA	8.302	5.243	2.448	7.459	8.800	0

Fig. 1. Principal coordinates for MMD analyses.
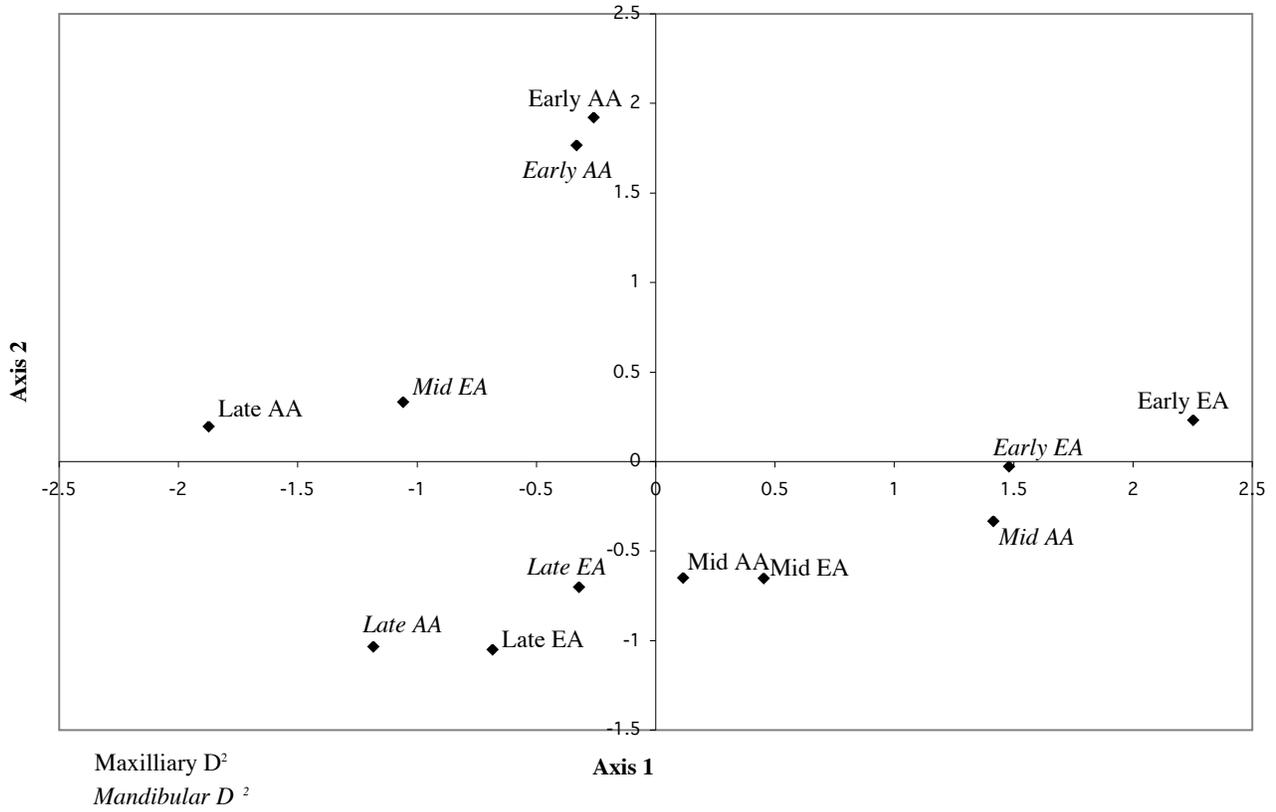
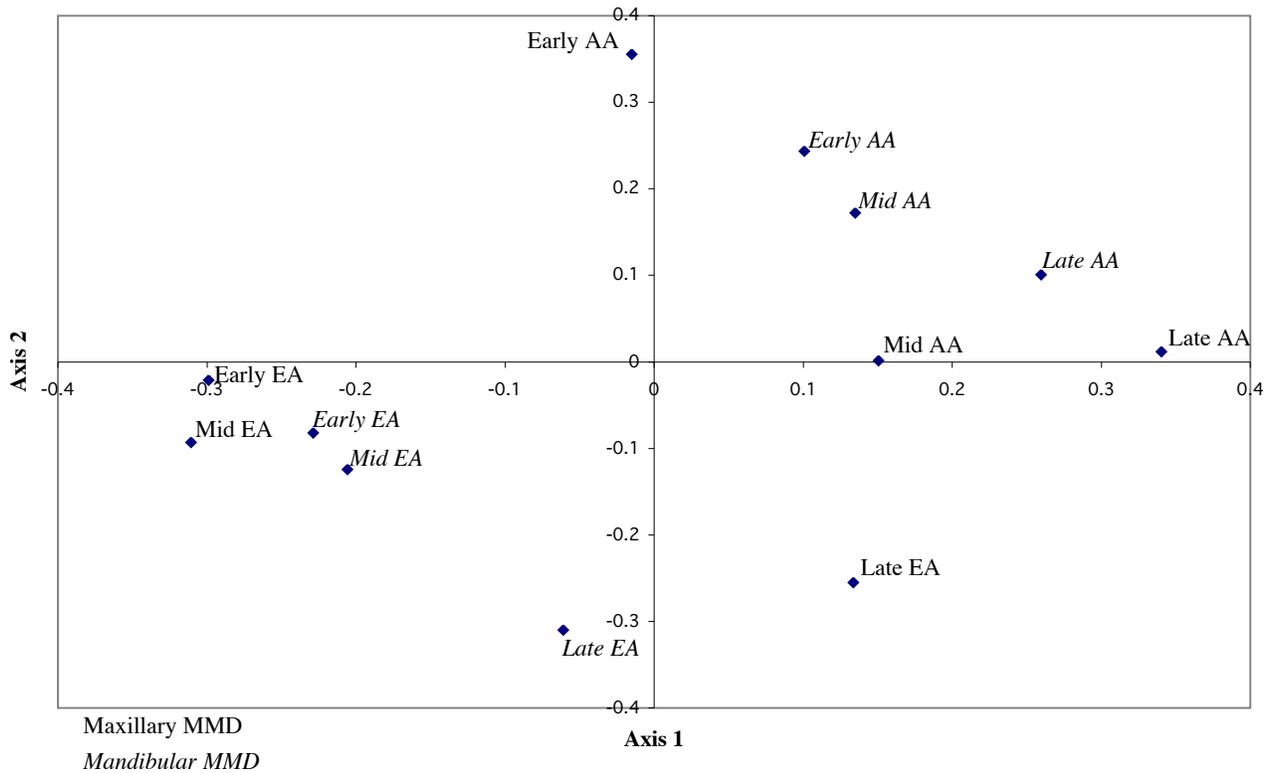**Fig. 2**. Principal coordinates for D² analyses.

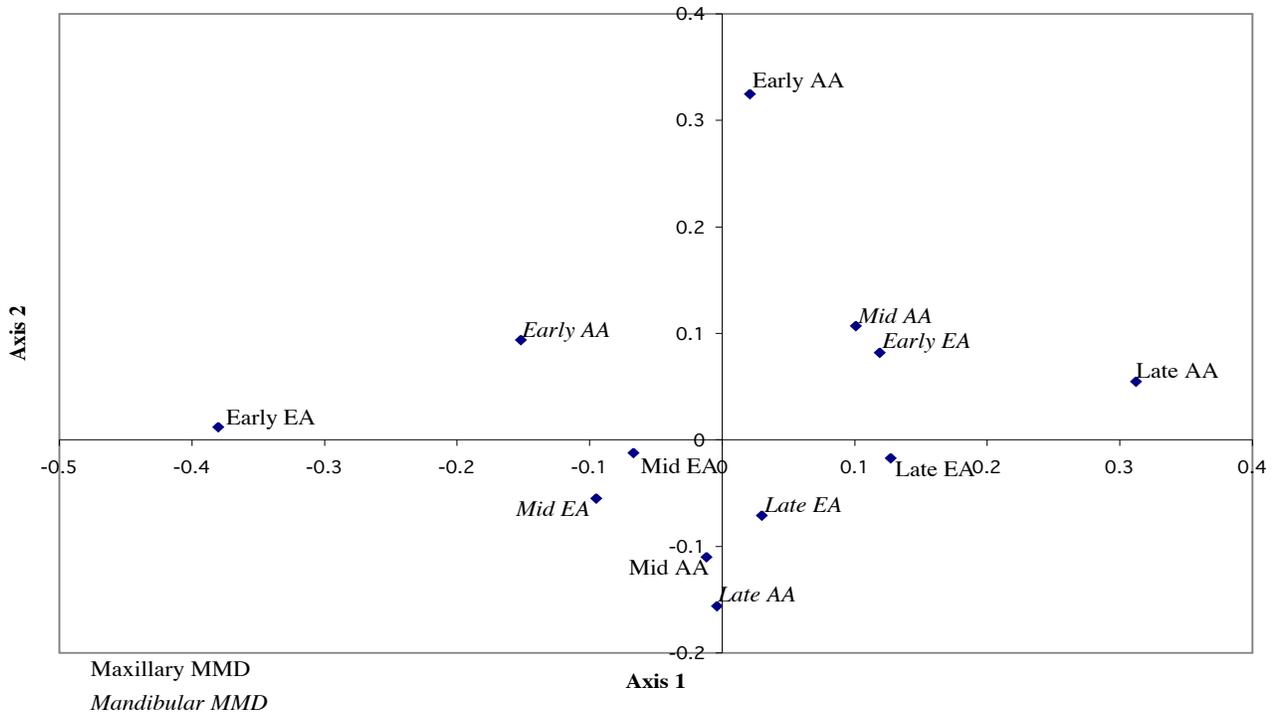**Fig. 3**. MMD Principal coordinates after procrustes transformation.

Fig. 4. D² Principal coordinates after procrustes transformation.
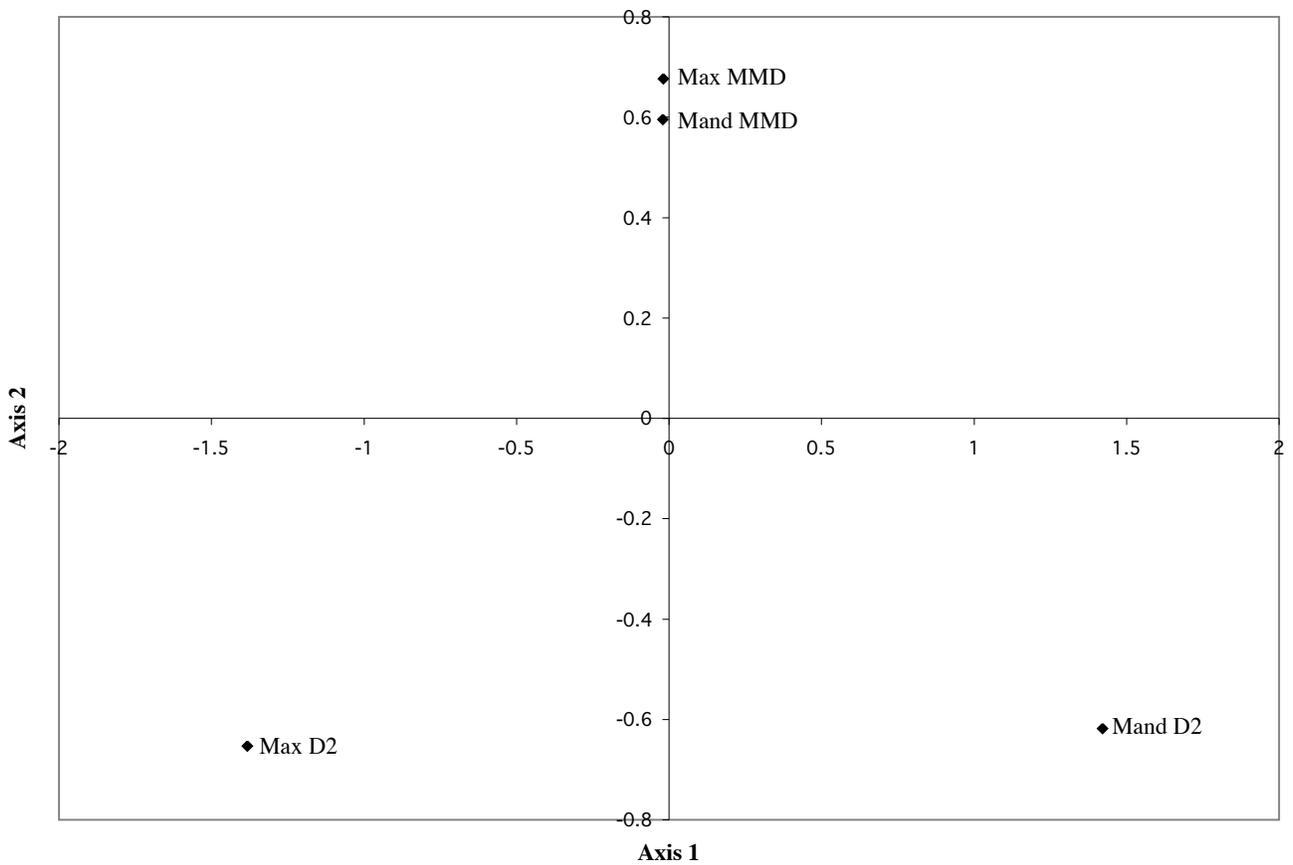

Fig. 5. Principal coordinates of residuals.

and Figure 2, which shows the same relationships for D^2 analyses.

Procrustes analysis

Figures 3 and 4 show the relationships between the six samples after rotation and scaling of the principal coordinates for MMD and D^2, respectively. The coordinates for maxillary MMD results acting as a baseline for both tables. Each of the other groups has been redrawn to its best fit, meaning the one that yields the smallest residual. The residuals between all the groups are summarized in Table 7. There is no test of significance for R^2, but it can be seen that all the values are relatively small except for between the D^2 for maxillary and mandibular characteristics. It is possible to simplify this table by performing a principal coordinates analysis for this R^2 matrix and display the relationships in the simplest geometric space. A graph of these coordinates shows relationship between the four methods of determining affinity. Figure 5 shows that the two MMD matrices are in nearly perfect agreement. The two D^2 matrices are quite different from each other, but neither is more different from the MMD matrices than the other.

It remains to be explained why the D^2 matrices are so different from each other. One possible explanation is a lack of differences between the samples being studied in these particular traits. In fact, among the traits used for the mandibular D^2 analysis, there is half the average difference in expression between groups as there is in the maxillary D^2 and MMD, and one quarter as much difference as in mandibular MMD.

CONCLUSIONS

Overall, there is very good agreement between the biological distance matrices generated using MMD and pseudo-Mahalanobis' D^2 statistics. Both statistics have their place in the analysis of biological distance, especially when utilizing characteristics of dental morphology. As with all statistics, the MMD and D^2 are limited by the data they analyze. If there is little difference between samples for the characteristics in question, the results will show small distances; if the differences are large for those particular characteristics, the distances will be large as well. A careful evaluation of the data should be made before attempting any measure of affinity.

When there are many traits available for analysis and they have little inter-trait correlation, MMD is appropriate. When the data consist of a relatively few, correlated traits, a pseudo-Mahalanobis' D^2 is more accurately applied, as it makes no assumption about a lack of correlation between traits. In a large study, the use of both statistics may allow analysis of more of the collected data. If all things are equal and either statistic is applicable, MMD is simpler to use and more widely

comparable.

LITERATURE CITED

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

Berry AC, Berry RJ, Ucko PJ. 1967. Genetical change in Ancient Egypt. Man n.s. 2:551-568.

Brown MB. 1977. The tetrachoric correlation and its asymptotic standard error. Appl Statist 26:343-351.

Burnaby TP. 1966. Growth invariant discriminant functions and generalized distances. Biometrics 22:96-210.

Davis FJ. 1991. Who is Black? One nation's definition. University Park, PA: Pennsylvania State University Press.

de Souza P, Houghton P. 1977. The mean measure of divergence and the use of non-metric data in the estimation of biological distances. J Arch Sci 4:163-169.

Falconer DS. 1981. Introduction to quantitative genetics. New York : Longman.

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

Goodall CR. 1991. Procrustes methods and the statistical analysis of shape (with discussion). J Roy Stat Soc B 53:285-340.

Gower JC. 1971. Statistical methods of comparing different multivariate analyses of the same data. In: Hodson FR, Kendall DG, Tautu P, editors. Mathematics in the archaeological and historical sciences. Edinburgh: Edinburgh University Press.

Gower JC. 1975. Generalized Procrustes analysis. Psychometrika 40:33-51.

Grewal MS. 1962. The rate of genetic divergence in the C57BL strain of mice. Genet Res 3:226-237.

Haeussler AM, Irish JD, Morris DH, Turner CG II. 1989. Morphological and metrical comparison of San and central Sotho dentition from southern Africa. Am J Phys Anthropol 78:115-122.

Irish JD. 1997. Characteristic high- and low-frequency dental traits in sub-Saharan African populations. Am J Phys Anthropol 102:455-467.

Irish JD. 1993. Biological Affinities of Late Pleistocene through Modern African Aboriginal Populations: The Dental Evidence. Ph.D. dissertation, Arizona State University.

Irish JD, Turner CG. 1990. West African dental affinity of late Pleistocene Nubians: peopling of the Eurafrican-

South Asian triangle II. Homo 41:42-53.

Jackson DA. 1995. PROTEST: a PROcrustean Randomization TEST of community environment concordance. Ecoscience 2:297-303.

Konigsberg LW. 1990. Analysis of prehistoric biological variation under a model of isolation by geographic and temporal distance. Hum Biol 62:49-70.

Mahalanobis PC. 1936. On the generalized distance in statistics. Proc Nat Inst Sci India 12:49-55.

Manly BFJ. 1994. Multivariate statistical methods: a primer. London: Chapman and Hall.

Mizoguchi Y. 1977. Genetic variability in tooth crown characters: Analysis by the tetrachoric correlation method. Bull Nat Sci Mus, Series D 3:37-62.

SAS Institute Inc. 1990. SAS/STAT Users guide, Version 6, 4th ed. Cary, NC: SAS Institute, Inc.

Scott GR, Turner CG II. 1997. The Anthropology of modern human teeth: dental morphology and its variation in recent human populations. Cambridge: Cambridge University Press.

Turner CG, Nichol CR, Scott GR. 1991. Scoring procedures for key morphological traits of the permanent dentition: the Arizona State University dental anthropology system. In: M Kelley, Larsen CS, editors. Advances in dental anthropology. New York: Wiley-Liss, p 13-31.

Turner CG. 1987. Late Pleistocene and Holocene population history of East Asia based on dental variation. Am J Phys Anthropol 73:305-321.

13th International Symposium on Dental Morphology
First Announcement

The 13th International Symposium on Dental Morphology is taking place from Wednesday 24 to Saturday 27 August 2005, hosted by the University of Lódz, Poland. The conference web-site is at: http://www.biol.uni.lodz.pl/antropolog/conference/index.html

Files can be downloaded from the web-site for 1) Symposium Registration, 2) Presenter's Information, and 3) Guideline for manuscript preparation with presenter's instructions. Documents should be completed and return by the 28th February 2005.

Scientific Program: The Scientific Programme will be held in the conference facilities at the University of Lódz and will follow the general pattern of previous meetings, with single oral and poster sessions.

Abstracts: We welcome abstract submission, with the deadline of 28th February 2005. An abstract submission form and a presenter's form is available from the organizers, with the choice of preferred option of poster or oral communication.

Symposium Proceedings: The Symposium proceedings will consist of the presentations as short papers. Our proposed deadline for manuscripts will be 31th May 2005. See information on the web-site for submission formats. The Symposium and the accommodation are organized in the University Conference Centrum. The Centrum is set in the University District in very pleasant grounds, close to the city center (Piotrkowska street). We will be using all the conference facilities on site. The accommodation includes single and double rooms. Travel from this venue to our social events and return is included in the fee.

Symposium Costs:

Participant	130 Euro by 28.02.2005	180 Euro by 31.05.2005
Accompanying person	50 Euro by 28.02.2005	80 Euro by 31.05.2005

Please note: We regret that any cancellation after 01.07.2005 will not be refundable.

Conference Fee covers: book of abstracts, the Symposium Proceedings, attendance to all sessions, refreshments during the meeting, conference facilities, the Welcome Reception, sightseeing of Lódz, grill party and the Gala Dinner. Much more information is available on the website.

Editor's Note: This information is abstracted from a detailed e-mail sent in early October. Be certain to refer to the web-site for specifics.