AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 139:68–76 (2009)

Understanding Race and Human Variation: Why Forensic Anthropologists are Good at Identifying Race

Stephen Ousley,[1]* Richard Jantz,[2] and Donna Freid[2]

[1]*Department of Applied Forensic Sciences, Mercyhurst College, Erie, PA*
[2]*Department of Anthropology, The University of Tennessee, Knoxville, TN*

KEY WORDS race; craniometric variation; discriminant function analysis; forensic anthropology; classification

ABSTRACT American forensic anthropologists uncritically accepted the biological race concept from classic physical anthropology and applied it to methods of human identification. Why and how the biological race concept might work in forensic anthropology was contemplated by Sauer (Soc Sci Med 34 [1992] 107–111), who hypothesized that American forensic anthropologists are good at what they do because of a concordance between social race and skeletal morphology in American whites and blacks. However, Sauer also stressed that this concordance did not validate the classic biological race concept of physical anthropology that there are a relatively small number of discrete types of human beings. Results from Howells (Papers of the Peabody Museum of Archaeology and Ethnology 67 [1973] 1–259; Papers of the Peabody Museum of Archaeology and Ethnology 79 [1989] 1–189; Papers of the Peabody Museum of Archaeology and Ethnology 82 [1995] 1–108) and others using craniometric and molecular data show strong geographic patterning of human variation despite overlap in their distributions. However, Williams et al. (Curr Anthropol 46 [2005] 340–346) concluded that skeletal morphology cannot be used to accurately classify individuals. Williams et al. cited additional support from Lewontin (Evol Biol 6 [1972] 381–398), who analyzed classic genetic markers. In this study, multivariate analyses of craniometric data support Sauer's hypothesis that there are morphological differences between American whites and blacks. We also confirm significant geographic patterning in human variation but also find differences among groups within continents. As a result, if biological races are defined by uniqueness, then there are a very large number of biological races that can be defined, contradicting the classic biological race concept of physical anthropology. Further, our results show that humans can be accurately classified into geographic origin using craniometrics even though there is overlap among groups. Am J Phys Anthropol 139:68–76, 2009. ©2009 Wiley-Liss, Inc.

Forensic anthropology is most often employed in the personal identification of human remains from crime scenes or mass disasters. Part of the identification process in identifying unknown remains is the construction of the biological profile, with parameters such as age, race, sex, and stature to compare to possible individual identifications. The continued use of race in forensic anthropology has been criticized because of the recent emphasis in biological anthropology to disprove the biological race concept of classic physical anthropology when discussing human variation. Indeed, many contemporary textbooks in forensic anthropology structure human variation in terms of three main races, stocks, or ancestral groups (Bass, 2005; Byers, 2005; Klepinger, 2006). Although a shift in terminology has been underway in forensic anthropology, with "ancestry" used more often in place of "race," in many case reports the classic physical anthropology terms such as "Caucasoid," "Mongoloid," or "Negroid" are still seen.

Unfortunately, the frequently ambiguous use of "race" in publications has led to many misunderstandings. This ambiguity is also reflected in the pages of the *American Journal of Physical Anthropology* in an article titled *"Race" Specificity and the Femur/Stature Ratio* (Feldesman and Fountain, 1996), in which race is referred to repeatedly in quotation marks, but never defined or explained. In this article, "race" will be used in its normal American sense, to refer to social race, and "biological race" will be used for the biological sense of the word. A reasonable but subjective definition of a biological race comes from Brues' definition (1977, p 1), "a division of a species which differs from other divisions by the frequency with which certain hereditary traits appear among its members.", which parallels definitions from Boyd (1950) and Hooton (1926). Thus, biological races in humans as well as animals are supposed to share heritable traits that make them similar to each other and also make them distinct from other biological races. In zoology, one statistical approach to discerning subspecies was the "75% rule" (Amadon, 1949) of separation as a criterion for taxonomists using morphological traits.

Sauer (1992) recognized the theoretical tension between forensic and biological anthropology in his article with the subtitle "If races don't exist, why are forensic anthropologists so good at identifying them?" While no accuracy figures were given, Sauer (1992) concluded that forensic anthropologists were good at identifying races because there is a concordance between American social races and skeletal biology, specifically, cranial morphology in black and white Americans. These two groups are the most

---

*Correspondence to: Stephen Ousley, Department of Applied Forensic Sciences, Mercyhurst College, Erie, PA.
E-mail: sousley@mercyhurst.edu

Received 28 April 2008; accepted 26 November 2008

DOI 10.1002/ajpa.21006

likely to have historically required forensic identification in most areas of the US. Sauer (1992) maintained that in the US, people of African ancestry were likely to have a different morphology from those with European ancestry. However, Sauer (1992) also concluded that the ability of forensic anthropologists to classify individuals does not validate the classic biological races from physical anthropology in the broader sense, i.e., that humans form a small number of discrete entities that are inherently different from each other.

Explicit tests of Sauer's (1992) hypotheses have not been published until now, though Goodman (1997) challenged the view that forensic anthropologists can accurately identify race at all, citing four cases in which forensic anthropologists were incorrect in their assessment of race. Goodman (1997, p 22) concluded that "At best, in other words, racial identifications are depressingly inaccurate. At worst, they are completely haphazard". Four misjudgments, compared to what must be many thousands of cases in which forensic anthropologists have been correct, do not make a compelling argument. Additionally, Armelagos and Goodman (1998, p 370) maintained "The use of race in forensic research has probably led to countless misidentifications." Many historical and theoretical reasons have been provided for why there should be no association between social and biological race in the US (Goodman and Armelagos, 1996; Williams et al., 2005). There is genetic evidence of up to ~20% European admixture in some African-Americans communities, which would make the two groups more similar (Parra et al., 1998). Some racial definitions in the US have depended on the One-Drop Rule, whereby "one drop" of African ancestry would qualify a person as black. In fact, there does not appear to be a consistent legal definition of what "race" means (Wright, 1995). Race definitions have changed over time, and in fact at one time the Irish were not considered part of the white race for immigration purposes. Finally, human variation is supposed to show a clinal pattern with no distinct boundaries, and Livingstone's (1962, p 279) quote is often cited: "There are no races, there are only clines." Many have also repeated the claim that the traits that supposedly define biological races are inherited independently and do not form distinctive trait clusters by which one could objectively define biological races. These findings have changed how physical anthropology is taught and has resulted in the frequently heard mantra "Race doesn't exist" (Lieberman and Kirk, 2004).

On the other hand, social race has greatly influenced mating in the US, reflecting positive assortative mating and limiting gene flow among groups. Up to 1970, the black-white interracial marriage rate for whites was ~0.1% and for blacks was 1%, based on US Census data (Fryer, 2007). The interracial marriage rate has increased since then but rates were still relatively low based on census data from 2000, with a rate of ~0.3% for whites and 4% for blacks. A historically low rate of interracial marriage in the US should come as no surprise when racism, especially institutional racism, has been prevalent. The first colonies to enact antimiscegenation laws were Maryland and Virginia, with official penalties for interracial marriage that included banishment and jail. In fact, laws against interracial marriage in 16 states including Virginia were not repealed until after the 1967 Supreme Court decision in *Loving v. Virginia* (Fryer, 2007). Unofficial social penalties for interracial relationships and marriage included violence and murder. Marriage is a very public social declaration, so marriage rates from census data will not reflect all interracial relationships, and personal ads may better represent human behavior in such relationships. In a recent study of whites who placed online dating ads, ~50% said that race was not important, but 90% of those individuals replied only to white respondents (Hitsch et al., 2004).

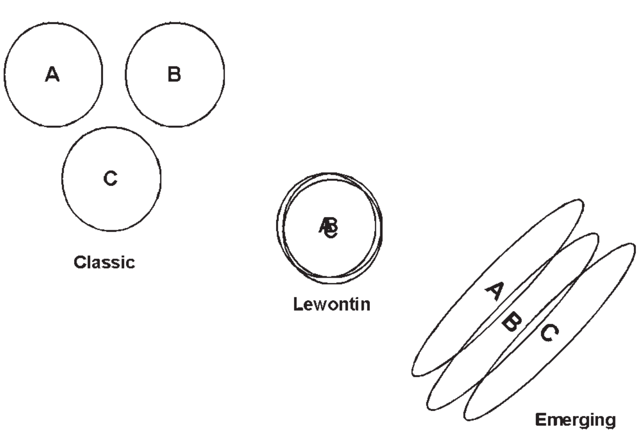

Fig. 1. Illustration of how views of human variation have changed, beginning with the classic typological view of physical anthropology. Lewontin's view has dominated for more than 30 years, but the emerging view of human variation takes into account the covariation of variables. Note that the emerging view recognizes large amounts of within-group variation compared to among-group variation, yet also allows separation among groups.

In examining human genetic variation on a worldwide scale, Lewontin's (1972) study of human variation using classic genetic markers has been cited as evidence that differences among human groups are too small to allow accurate classification. Lewontin estimated that ~85% of human genetic variation is found within populations, ~8% is found within populations of the same race or regional grouping, and only 6% is found among races or regions. Lewontin analyzed each of the genetic markers independently and overlooked the fact that some markers are significantly correlated with others and therefore not independently distributed among groups. Edwards (2003) confirmed that Lewontin's findings are correct at the single-locus level, meaning that single loci will show great overlap among groups, but analyzing multiple loci will produce less overlap among groups and reveal a more realistic picture of among-group variation. Additionally, Lewontin's conclusions seem more likely to be anomalous after the publication of numerous molecular analyses utilizing combinations of SNPs, STRPs, VNTRs, Alu insertions, and other molecular features that indicate strong geographic patterning in worldwide samples and accurate classification of groups (Pritchard et al., 2000; Rosenberg et al., 2002; Bamshad et al., 2003; Allocco et al., 2007) despite large amounts of within-region variation (Jorde and Wooding, 2004). The evolution of how scientists have viewed human variation is shown in Figure 1. The emerging view shown in Figure 1 also illustrates how accurate classification is possible despite large amounts of within-group variation.

In examining human craniometric variation on a worldwide scale, studies of the Howells craniometric data have produced consistent results. Relethford (1994, 2002) found worldwide levels of craniometric variation in the Howells data on a par with Lewontin's estimates. In contrast to Lewontin, Howells (1973, 1989) and Roseman and Weaver

(2004) found strong geographic patterning in the same data, and this patterning is present at an early age (Vidarsdottir et al., 2002). Other studies have used discriminant function analysis (DFA) to classify one individual at a time from known samples into Howells' groups, and their results seem to seem in agreement with those of Lewontin (1972). Ubelaker et al. (2002) classified 50 individuals from a likely 16th or 17th Spanish cemetery, and while most classified into the geographically closest groups, from Austria, Egypt, Hungary, and Norway, a good number classified into the Howells groups from Asia. Ubelaker et al. (2002), echoing Ousley and Jantz (1996), concluded that DFA should be used with caution when classifying populations that are not represented in the reference populations. Williams et al. (2005) classified 42 Nubians into the Howells groups with the expectation that all Nubians would classify into Howells' Egyptian group because the Egyptians are the closest group temporally and geographically. However, some of their Nubians were classified into groups as far away as Japan, Australia, and the New World. The main conclusion of Williams et al. (2005) was that classification methods cannot work because human variation is very limited and craniometric affinities of groups do not reflect geography: "The possibility that skeletal material could be sorted by geographic origin, at any other level than geographic extremes, is quite small" (Williams et al., 2005, p 345).

Several problems are apparent in the approach, results, and conclusions of Williams et al. (2005). Ten Nubians in their sample showed typicality probabilities that were too low ($P < 0.05$) to be assigned with confidence. In fact, eight of the ten, or 19% of the total sample, showed a typicality probability for the group they classified into of 0.005 or less, as seen in their Table 1. When individuals show such low typicality probabilities, they are outliers, and measurement or data entry error should be checked for first (Maindonald and Braun, 2003; Hair et al., 2006; Tabachnick and Fidell, 2007). As has been pointed out, DFA will classify any and all measurements and individuals, whether or not the measurements are correct, even if the measurements come from another species or a soccer ball (Freid et al., 2005; Ousley et al., 2007). Also, there must be something wrong with their measurements because the 12 measurements they specified include two measurements that Howells never collected: palate length and minimum frontal breadth. If they mistakenly entered palate length as palate breadth, or minimum frontal breadth as frontomalar breadth, that would explain the low typicality probabilities. Otherwise, Williams et al. (2005) performed their analyses using only 10 variables. However, there is no way of identifying which measurements they used or how they used them, because the authors have refused repeated requests for their Nubian data or Fordisc results from several researchers.

Despite some disagreements in interpretation, assessing human variation using craniometrics and multivariate methods is the best way to test Sauer's (1992) conclusions for several reasons. First, craniometrics reflect aspects of cranial morphology suggested by Sauer (1992) and can be easily analyzed using several multivariate statistical methods that allow more powerful tests of variation. Using multiple measurements provides a better overall morphological assessment of variation, and avoids problems with using only a few measurements. Second, multivariate classifications of craniometrics within traditional races have found significant variability, such as in American whites (Ousley and Jantz, 2002), African groups (Spradley, 2006; Spradley et al., 2008b), Hispanic groups (Ross et al., 2004; Slice and Ross, 2004; Ross et al., 2005; Spradley et al., 2008a), Native Americans (Ousley and Billeck, 2001; Ousley et al., 2005), and East Asian groups (Ousley et al., 2003). Third, while craniometrics show an association with environmental factors such as mean temperature (Beals et al., 1984), they and other measurements have been shown to reflect genetic relationships in animals with known pedigrees, including humans (Cheverud, 1988; Konigsberg and Ousley, 1995), and thus qualify as heritable traits in identifying biological races following Brues' (1977) definition. Finally, craniometric data sets with numerous measurements and large sample sizes are available from modern Americans as well as populations from around the world. This article will scientifically evaluate the conclusions of Sauer (1992) using modern, historic, and prehistoric craniometric data. It will also explore the apparently contradictory results from examining group affinities and individual classifications.

TABLE 1. Howells male groups used in worldwide comparisons and their geographic region

Group name	*N*	Abbreviation	Continent/Region
Ainu	48	AINM	East Asia
Andaman Islanders	35	ANDM	East Asia
Anyang	42	ANYM	East Asia
Arikara	42	ARIM	America
Atayal	29	ATAM	East Asia
Australian	52	AUSM	SW Pacific
Berg	56	BERM	Europe
Buriat	55	BURM	East Asia
Bushman	41	BUSM	Africa
Dogon	47	DOGM	Africa
Easter Islanders	49	EASM	Polynesia
Egypt	58	EGYM	Africa
Eskimo	53	ESKM	America
Guam	30	GUAM	SW Pacific
Hainan	45	HAIM	East Asia
Mokapu	51	MOKM	Polynesia
Moriori	57	MORM	Polynesia
Norse	55	NJAM	Europe
North Japan	55	NORM	East Asia
Peru	55	PERM	America
Philippines	50	PHIM	East Asia
Santa Cruz	51	SANM	America
South Japan	50	SJAM	East Asia
Tasmanians	45	TASM	SW Pacific
Teita	33	TEIM	Africa
Tolai	56	TOLM	SW Pacific
Zalavar	53	ZALM	Europe
Zulu	55	ZULM	Africa

N, number in each sample.

MATERIALS AND METHODS

Craniometric data from 353 individuals in the FDB were used to compare variation in white and black Americans. All were born in the 20th century and are of self-identified race and sex. Craniometric data collected by W.W. Howells were used in comparing groups from around the world, and measurement abbreviations follow Howells (1973). The Howells database consists of 2,504 individuals from 28 groups of males and 26 groups of females from around the world and from various time periods (Howells, 1996). The names of the groups, abbreviations, and region are listed in Table 1.

Because multiple variables provide a better assessment of overall morphological variation, several multivariate statistical methods were used, each with different advantages and assumptions. Discriminant function analysis (DFA) maximizes the differences among groups, so it provides a best case classification method if within-group variation is similar, but exaggerates underlying differences among groups. K-nearest neighbor (KNN) analysis relies on interindividual similarities rather than group similarities, but still relies on within-group variation in the original groups. Both of these classification methods record classification error rates for each group. The error rate is important because a classification procedure is best judged by how well it classifies known reference groups. Correct classification rates that are little better than random mean that there is no appreciable intergroup variation in the variables used. Accordingly, classification rates at a far greater rate than expected based on random allocation will be considered as support for the hypothesis that differences exist among groups. Cluster analysis is a more conservative test for group differences because individuals are naively classified into a specified number of natural groupings, and only natural group parameters are used. Additionally, principal components analysis (PCA) was employed in various analyses. The first principal component contains the greatest amount of variation present in all original measurements, and subsequent principal components represent progressively smaller amounts of variation. Often, the bulk of variation in a large number of measurements is expressed in far fewer principal components (Tabachnick and Fidell, 2007).

DFA uses multivariate methods developed over 70 years ago (Fisher, 1936; Mahalanobis, 1936) to classify individuals into the group they are most similar to using group means and the pooled within-group variance-covariance matrix (Huberty, 1994; Huberty and Olejnik, 2006; Tabachnick and Fidell, 2007). Additionally, an individual's posterior and typicality probabilities are calculated. Stepwise variable selection is a technique to identify the measurements that separate groups best. Fordisc 3.0 (Jantz and Ousley, 2005) and Systat (Systat Software Inc., 2004) were used to perform DFA and stepwise variable selection. We report classification percentages using the most often recommended way of estimating classification error rates, leave-one-out cross-validation. In this method, each individual is sequentially removed from the DFA, a function based on the rest of the sample is calculated, and the classification of the individual is recorded. The estimated error rate using leave-one-out cross-validation is not biased upwards and will better reflect error rates when applied to new cases (Huberty, 1994).

KNN analysis using a custom computer program and SAS (SAS Institute, 2001) was also used for classification. Unlike DFA, group membership and group means are not incorporated into the procedure. Instead, individuals are classified based on their similarity to other individuals. Multivariate distances are calculated to individuals, rather than to groups, and the most similar K individuals form the basis for classification (SAS Institute, 2001). KNN analysis using craniometrics is the basis for CRANID (Wright, 1992) and has been used to classify individuals, including an Egyptian mummy (Hughes et al., 2005). In these analyses, group assignment was based on $K = 1$, the single nearest neighbor.

K-means cluster analysis was also used to classify individuals using Systat. In contrast to the other procedures, it finds a specified number (K) of natural groups of individuals in a sample. At the beginning of the process, all members are placed in one group and the means for each variable are calculated. The member of the group that is most different from the grand mean is chosen as the seed for a second group. New means are calculated for each group, and each individual is then assessed as to which cluster it is most similar to, and the individual closest to a different cluster is then transferred to that cluster. Cluster means are recalculated whenever membership changes, and the process is repeated numerous times. In the process, cluster members may later be rejoined to their former cluster. At the end of the process, there are K clusters with minimized within-cluster variation and maximized among-cluster variation (Systat Software Inc., 2004; Tabachnick and Fidell, 2007).

TABLE 2. Results of K-means cluster analysis performed on 375 American black and white males from the FDB using a two-cluster solution

	Cluster 1	Cluster 2	
BM	17	132	89% in 2
WM	194	32	86% in 1
	211 (92% WM)	164 (80% BM)	

$$\frac{\text{Between Cluster SS}}{\text{Total SS}} = \frac{21,824}{191,079} = 11\%.$$

RESULTS

American Whites and Blacks

DFA using just two variables, basion-nasion length (BNL) and basion-prosthion length (BPL), separates American blacks and whites about 80% correctly, and using more variables improves classification accuracy (Jantz and Ousley, 2005). A discriminant function using 19 measurements magnifies the differences and can classify the same samples into social race 97% correctly. Using stepwise variable selection, only seven variables (BNL, BPL, biauricular breadth (AUB), nasal breadth (NLB), palate breadth (MAB), orbital height (OBH), in order of selection) are necessary to classify blacks and whites 95% correctly, and these variables are ones that can be visually appreciated by forensic anthropologists. BNL and BPL express relative prognathism, BBH is an expression of vault height, AUB is a measure of vault breadth, NLB of nasal breadth, MAB of palate breadth, and OBH, orbital height, representing orbital shape. Nearly all of these measurements represent morphological configurations mentioned in forensic anthropology texts as valuable in determining race visually: prognathism, the cranial index, nasal breadth, and orbit shape. Because DFA magnifies differences among groups, quantifying group differences using PCA will produce a better baseline measure of differences. When PCA was run on 19 basic measurements from the total sample of 375 black and white males, the first principal component, which comprises the greatest interindividual variation, separated black and white males 81% correctly. Further, in a K-means cluster analysis using the same 19 basic variables, 92% of cluster one members were white males, and 80% of cluster two members were black males; 89% of black males were placed into cluster two and 86% of white males were placed into cluster one (Table 2). Between-cluster variation was 11% of the total variation. These results indicate that there are significant differences between the two groups before being magnified by

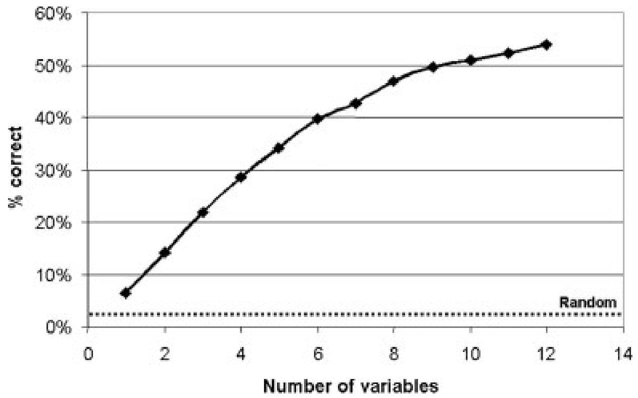

Fig. 2. Number of variables in DFA and classification accuracy for 27 Howells male groups. The mean percentage correct for two or more variables is significantly greater than chance at $P < 0.01$.

TABLE 3. Correct classification percentages of 27 Howells male groups using different combinations of variables

N vars	% Group correct	% Region correct
10	50	70
10 SW	56	75
15 SW	64	83
24 SW	74	89

The first entry for 10 variables includes the 10 variables used by Williams et al. (2005). SW, stepwise-selected measurements.

DFA, and support Sauer's (1992) assertion that there are morphological differences between American blacks and whites that can be visually appreciated. In other words, for these groups, there is a strong concordance between social race and biological differences.

Worldwide craniometric variation and classification

The other conclusions of Sauer (1992) and Williams et al. (2005) were tested by investigating patterns of worldwide craniometric classification. First, the effect of the number of variables on classification accuracy in the Howells groups was investigated (see Fig. 2), with higher classification accuracy resulting from a greater number of variables used. When using at least two variables, the mean classification accuracy in groups was greater than random (3.6%) at $P < 0.01$, and using more variables narrowed the variation in mean classification percentage by improving the lowest correct percentages. When the Howells groups were classified with the same 10 variables used by Williams et al. (2005) in a 28 group function, the 10 variables classified individuals 50% correctly on average, and 70% were assigned into a group from the same continent or region (Table 3). Using 10 stepwise-selected variables produced somewhat higher accuracies, with the lowest group accuracy at 13%, but the greatest performance was seen in using 24 stepwise-selected variables, with almost 75% of the individuals correctly classified into their own group and almost 90% classified into a group from the same region. Further, using 14 basic measurements in a KNN analysis classified individuals into their own group 31% correctly on average, far higher rate than expected by chance, and into the correct continent or region 56% correctly. Thus, individuals in the Howells data are more similar to other individuals from the same group than they are to individuals from other groups. The similarities and differences among groups are apparent whether looking at interindividual similarities (KNN analysis), or intergroup differences magnified by DFA, and there is clear regional patterning to their similarities. Occasionally some individuals were classified into groups from different continents in these analyses, but the number of those individuals so classified decreased as more variables were used. Using the 24 variable set, the only groups that were classified outside of their own continent involved the Egyptian and European samples. Five of 58 Egyptians (9%) classified into the Norse sample and 3 (7%) were classified into the Zalavar sample; 5 of 55 (9%) of the Norse sample were classified into Egypt and of 4 of 53 (8%) of the Zalavar sample were classified into Egypt. Egypt lies in the northeast corner of Africa, and these Egyptian-European results are in agreement with the consensus view that the Sahara Desert has been a more significant barrier to human groups than the Mediterranean Sea.

As a follow-up to the clear craniometric separation of American blacks and whites, a six-way DFA was performed on six of Howells' European and sub-Saharan African samples (Berg, Norse, Zalavar, Dogon, Teita, and Zulu). The function classified 82% of them correctly into their own group and 98% of them into a group from the same continent. When the groups were pooled into Europeans and Africans in a two-way function, similar to what Howells (1970) performed, 99% were correctly classified, and K-means cluster analysis revealed continental differences, with one cluster having 81% of the Africans and the other cluster having 91% of the Europeans. The groups from Europe and Africa would thus appear to meet Brues' (1977) definitions of a biological race in that they can be separated from each other very effectively using craniometrics.

However, within-continent analyses complicate the craniometric differences between continents. In Europe, DFA applied to Howells' Berg and Norse groups classified them about as well (83% correct) as DFA applied to American blacks and whites, and using the best variables improved the correct classification rate to 93%. Moreover, K-means cluster analysis separated 81% of the Berg and Norse samples into two different clusters. Also, a European three-way function for Berg, Norse, and Zalavar classified the groups 73% correctly, and the best variables classified them 85% correctly. Therefore, the three European samples would appear to meet Brues' (1977) definition of different biological races as well. In Japan, DFA using 18 variables classified Howells' Northern and Southern Japan samples 89% correctly, and K-means cluster analysis allocated 81% of each Japanese group into separate clusters. Therefore, the Northern and Southern Japan samples would also represent different biological races. It would seem that the number of biological races may be limited only by the number of samples, contradicting the classic view that there are only a few discrete biological races.

DISCUSSION

Our analyses of craniometric variation in black and white Americans using several multivariate statistical methods support Sauer's (1992) conclusion that objective morphological differences exist between American whites and blacks. We have demonstrated a concordance between social race and cranial morphology, at least in 20th century American blacks and whites. Other skeletal studies have reached similar conclusions (Edgar, 2002; Konigsberg and Jantz, 2002; Ousley and McKeown, 2003). Craniometric differences between American blacks and whites have not diminished since the 19th century, though both groups have changed since then (Wescott and Jantz, 2005). The probable reasons for biological differences should be familiar to many. American blacks and whites originated from different continents, and American blacks are largely composed of West African groups transported to the US for the slave trade. Europeans and Africans had been separated and experiencing different evolutionary forces for tens of thousands of years before migrating to the US. The high accuracy of the two-way DFA between the pooled Howells European and sub-Saharan African groups indicates they likely had differentiated. As mentioned, institutional racism and assortative mating within social race has prevented significant gene flow between them, which would make them more similar.

Part of the reason for the disagreement between forensic and biological anthropologists has been in their different approaches and goals. Forensic anthropologists answer practical questions of age, sex, and race to construct the biological profile and narrow down possible identifications. In examining American blacks and whites, forensic anthropologists would naturally think in terms of two biological races because of the concordance between social and morphological race. Identifying social race, available in missing persons reports, would be the stopping point. Biological anthropologists would explore within-group variation further. These findings illustrate the essential difference between a forensic analysis and a biological analysis: forensic analysis produces practical information useful for forensic identification, while a biological analysis provides insight about relationships among arbitrarily defined populations, which may be defined by social races, breeding populations, language, nationality, time periods, and other criteria.

Sauer's (1992) additional suggestion that differences in American blacks and whites did not validate the traditional biological race concept is likewise supported by our results. On a worldwide scale, humans show geographically patterned variation when classified as groups and individuals, and although there is a good deal of overlap between groups and much variation within groups, individuals and groups can nonetheless be classified at a rate far greater than chance on the group and regional level. The classifications of the Howells data echo Howells' (1973, 1989) results that show rather strong geographic patterning, and there is clearly enough craniometric variation among groups to classify at rates far higher than random allocation. These findings directly contradict the conclusions of Williams et al. (2005) because individual crania are far more often than not classified into the group they are part of, or into a group from the same region. The classification rates are not 100% because of overlap among groups, consistent with other studies (Howells, 1970, 1989, 1995; Relethford, 1994, 2002; Roseman, 2004; Roseman and Weaver, 2004), and contradicting the biological race concept of classic physical anthropology.

In the Ubelaker et al. (2002) study, most individuals classified into the geographically closest Howells groups, from Egypt or Europe. As Ousley and Jantz (1996) point out, when classifying individuals that are not represented in the reference populations, caution is warranted. DFA will indicate the group that an individual is closest to morphologically, and should not be interpreted as a literal and binding classification. Also, individuals from countries like Spain that represented a world empire may well be morphologically heterogeneous, as Ubelaker et al. (2002) had noticed before their metric analysis. The Iberian Union (1580–1640) of the kingdoms of Castile, Aragon, and Portugal included parts of the Mediterranean, the Americas, coastal areas of Africa, India, Indonesia, the Philippines, Japan, and Guam, and Iberia had been part of Arab and Moor empires that stretched across the Mediterranean for hundreds of years. The morphological diversity of the Ubelaker et al. (2002) Spanish individuals may well reflect their heterogeneous origins, as is reflected in molecular studies (Casas et al., 2006; Alvarez et al., 2007). However, the Spanish centroid—the mean Spanish morphology—likely shows greatest similarity to the Howells European and Egyptian centroids.

In our classifications of the Howells data, some individuals were classified into groups from different continents, but those classifications largely disappeared when more variables were used. If Williams et al. (2005) used the correct measurements, they analyzed the Nubians in a 28-way function using 10 variables and maintained that a classification rate of less than 100% into Howells' Egyptians was evidence of failure. Thus, their null hypothesis was that of an extreme typologist: that all groups are expected to be unique with no overlap among groups. By the middle of the 20th century, many physical anthropologists had already acknowledged overlap among groups, though some still argued that a relatively small number of human races existed (e.g. Coon, 1965). More recent statistical comparisons among Howells' groups as well as modern forensic groups show considerable overlap and less than perfect classification, even when using more variables (Howells, 1995; Ousley and Jantz, 1997). In this case, a correct classification rate of 3.6% (1/28) would be expected by chance. As a matter of fact, even with strong indications of measurement error, the most common classification, eight of 32, or 25% of their Nubian sample, was into Howells' Egyptian group. It is also important to note that Williams et al. (2005) formed their conclusions about group similarities based on compiling individual classifications. Comparing group centroids is the best way to compare group relationships. Based on the individual results of Williams et al., their Nubian sample centroid is most likely closest to Egypt.

Other results indicate that the conclusions of Williams et al. (2005) are erroneous. Freid and Jantz (2005) analyzed Nubian craniometric data (93 females, 144 males) from the fortress at Mirgissa that had been collected during the UNESCO sponsored excavations (Vercoutter, 1976). When each individual was classified into the Howells groups using 17 variables in a 28-way function, 142 out of 237 (60%) were closest to Howells' Egyptian sample, and 183 out of 237 (77%) were closest to one of Howells' African groups. These results are consistent with what Williams et al. (2005) maintained was to be expected, because the Egyptian group is the closest geographically. Geography is often a proxy for population history, because groups that are closer to each other have more likely exchanged more genes directly or through other nearby groups simply because of proximity. Therefore, migration, gene drift, and gene flow likely influence modern human craniometric variation more than selection alone because through them, morphological changes can occur at a far greater rate.

TABLE 4. Classifications using DFA with various groups

Groups in DF	Variables	% Correct	Why?
BM vs. WM	19	97	Biological race
BM vs. WM vs. CHM vs. NAM	25	96	Biological race
BM vs. WM vs. JM vs. NAM	25	84	Biological race
JM vs. CHM vs. VM	25	80	Geography
Arikara vs. Sioux Females	7	87	Tribe
Nagasaki vs. Tohoku Males	25	94	Geography
N Japan vs. S Japan Males	18	89	Geography
WM born 1840-1890 vs. WM born 1930-1980	10	96	Time

BM, American black males; CHM, Chinese males; JM, Japanese males; NAM, Native American males; VM, Vietnamese males; WM, American white males.

Why was biological race considered an explanation for human differences, and why does it remain so for some? The socially inherited concept of race no doubt shapes interpretations, but so do interpretations of any inherent differences among human groups. Examining variation in different combinations of groups reveals a confirmation bias for the variable that is used to define groups, most often biological race. Craniometric comparisons of various groups from Ousley and Jantz (2002) are shown in Table 4 and the first few examples may seem to support traditional racial divisions of mankind. In the first comparison, biological race seems to be the reason that white and black males are different, because we assume that race is the controlling variable, the primary difference between them. When Chinese and Native American groups are added, results are still consistent with the traditional race concept. But in the third example, if Japanese are substituted for Chinese, the accuracy decreases because black and Japanese males tend to misclassify as each other. Further classifications in Table 4 among groups traditionally considered part of the same biological race were also highly accurate. A three-way DFA using Japanese, Chinese, and Vietnamese males classifies them quite well, but the differences among them are in language and nationality. Females from two Native American tribes, Arikara and Sioux, can be classified quite accurately, and tribe or language defines each sample. Within Japan, DFA can differentiate between modern Japanese from the north (Tohoku) and south (Nagasaki) even better, and in this case the groups are defined by geography. These differences parallel those between the Howells North and South Japanese males. Finally, white males born between 1840 and 1890 can be separated from white males born 1930 to 1980 very well, and they are distinguished by time, and would appear to qualify as different races. In all of these analyses, the groups were categorized by a variable and differences were found. While race has been traditionally used to explain why the groups are different, time as an explanation may be more difficult to grasp. But time per se is not the reason the two groups are different. Time in this example is correlated with vast improvements in nutrition, medical care, and hygiene in the US, which have produced secular changes in the later population. Relaxed selection and gene flow from new immigrants may have also contributed to the changes. The northern-southern dichotomy seen in modern Japanese represents considerable variation within Japan in other biological systems as well. These examples demonstrate that though the group qualifiers change, the qualification is not directly related to why the groups are different. In the first two examples, race does not directly explain differences, just as language per se does not, nor does region, nor geography, nor distance, nor tribe, nor time. Instead, all of these comparisons involve differently defined populations with different origins or histories. Each of the defining variables is arbitrary but is related to differences in origins, histories, environments, and reproductive barriers. Groups separated through social mechanisms, language, geography, or time can differentiate due to genetic drift and other evolutionary forces, and those qualifiers were likely factors restricting gene flow among the groups.

CONCLUSIONS

The Howells craniometric data provide a rich data set for testing hypotheses about human variation. Another significant advantage to the Howells data is the large number of variables collected. As we demonstrated, the number of variables analyzed affects classification accuracy. There is an obvious parallel in examining one genetic system such as ABO blood group and drawing conclusions based on that single system. Lewontin's (1972) conclusions were likewise based on univariate frequencies from a few genetic systems. However, as we and others have shown, many measures of human variation are correlated, requiring multivariate methods. The Howells data also has no interobserver errors, which likely explain the anomalous results of Williams et al. (2005).

In investigating the connection between social race and biology, it is clear that race in the US is a social phenomenon with biological consequences due to positive assortative mating and institutional racism: whatever differences there were between ancestral groups from Europe and Africa were not obliterated because of very low historic gene flow between them in the US, despite theoretical and historical reasons why social races may not reflect biology. In this regard, race (i.e., the history of American race relations) helps explain modern craniometric variation in American blacks and whites.

Worldwide craniometric variation shows strong geographic patterning. However, if biological distinctiveness is an accepted criterion for biological races, a very large number of biological races can be discerned using craniometric data alone. Given this fact and the many populations with unique histories, it makes sense to collect data from as many populations as possible to aid in accurate classification, as Howells (1995) and Ubelaker et al. (2002) concluded. With other biological systems and traits, the distribution and number of biological races will change. There are so many possible distinctive biological races that the concept is virtually meaningless. We can only concur with Howells' (1995, p 103) modification of Livingstone's 1962 quote: "There are no races, only populations."

ACKNOWLEDGMENTS

The authors thank Alain Fremont for access to Vercoutter's Nubian craniometric data from Mirgissa. Hajime Ishida provided craniometric data from modern Japanese. At Mercyhurst College, E. Susanne Daly helped with historical research and Alexandra Klales assisted designing several figures. The authors express their thanks to symposium organizers, Heather Edgar and Keith Hunley, who provided much encouragement. They also thank the anonymous reviewers and the AJPA editors for their feedback.

LITERATURE CITED

Allocco DJ, Song Q, Gibbons GH, Ramoni MF, Kohane IS. 2007. Geography and genography: prediction of continental origin using randomly selected single nucleotide polymorphisms. BMC Genomics 8:68.

Alvarez JC, Johnson DL, Lorente JA, Martinez-Espin E, Martinez-Gonzalez LJ, Allard M, Wilson MR, Budowle B. 2007. Characterization of human control region sequences for Spanish individuals in a forensic mtDNA data set. Leg Med 9:293–304.

Amadon D. 1949. The seventy-five per cent rule for subspecies. Condor 51:250–258.

Armelagos G, Goodman AH. 1998. Race racism, and anthropology. In: Goodman AF, Leatherman TL, editors. Building a new biocultural synthesis: political-economic perspectives on human biology. Ann Arbor: University of Michigan Press. p 359–378.

Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA, Jorde LB. 2003. Human population genetic structure and inference of group membership. Am J Hum Genet 72:578–589.

Bass WM. 2005. Human osteology, 5th ed. Columbia, MO: Missouri Archaeological Society.

Beals KL, Smith CL, Dodd SM. 1984. Brain size, cranial morphology, climate and time machines. Curr Anthropol 25:301–330.

Boyd WC. 1950. Genetics and the races of man. Boston: Little, Brown and Company.

Brues A. 1977. People and races. New York: Macmillan.

Byers SN. 2005. Introduction to forensic anthropology, 2nd ed. New York: Pearson.

Casas MJ, Hagelberg E, Fregel R, Larruga JM, Gonzalez AM. 2006. Human mitochondrial DNA diversity in an archaeological site in al-Andalus: genetic impact of migrations from North Africa in medieval Spain. Am J Phys Anthropol 131:539–551.

Cheverud JM. 1988. A comparison of genetic and phenotypic correlations. Evolution 42:958–968.

Coon CS. The living races of man. New York: Alfred A. Knopf.

Edgar HJH. 2002. Biological distance and the African American dentition. Ph.D. dissertation. Ohio: Ohio State University.

Edwards AWF. 2003. Human genetic diversity: Lewontin's fallacy. Bioassays 25:798–801.

Feldesman MR, Fountain RL. 1996. "Race" specificity and the femur/stature ratio. Am J Phys Anthropol 100:207–224.

Fisher RA. 1936. The use of multiple measurements in taxonomic problems. Ann Eugenics 7:179–184.

Freid D, Jantz RL. 2005. Classification and evaluation of unusual individuals using Fordisc. Proc Am Acad Forensic Sci 11:293–294.

Freid D, Spradley MK, Jantz RL, Ousley SD. 2005. The truth is out there: how not to use Fordisc. Am J Phys Anthropol Suppl 40:103.

Fryer RG Jr. 2007. Guess who's been coming to dinner? Trends in interracial marriage over the 20th century. J Econ Perspect 21:71–90.

Goodman AH. 1997. Bred in the bone? Sciences 37:20–25.

Goodman AH, Armelagos GJ. 1996. The resurrection of race: the concept of race in physical anthropology in the 1990s. In: Reynolds LT, Lieberman L, editors. Race and other misadventures: essays in honor of Ashley Montagu in his ninetieth year. Dix Hills, New York: General Hall. p 174–186.

Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL. 2006. Multivariate data analysis, 6th ed. Upper Saddle River, NJ: Pearson.

Hitsch G, Hortacsu A, Ariely D. 2004. What makes you click? An empirical analysis of online dating. UCSC Economics Department Seminars Paper 3. Available at: http://repositories.cdlib.org/ucsc econ seminar/winter2005/3.

Hooton EA. 1926. Methods of racial analysis. Science 63:75–81.

Howells WW. 1970. Multivariate analysis for the identification of race from crania. In: Stewart TD, editor. Personal identification in mass disasters. Washington, DC: National Museum of Natural History, Smithsonian Institution. p 111–121.

Howells WW. 1973. Cranial variation in man. Papers of the Peabody Museum of Archaeology and Ethnology, Vol. 67. Cambridge, MA: Harvard University. p 1–259.

Howells WW. 1989. Skull shapes and the map. Craniometric analyses in the dispersion of modern Homo. Papers of the Peabody Museum of Archaeology and Ethnology, Vol. 79. Cambridge, MA: Harvard University. p 1–189.

Howells WW. 1995. Who's who in skulls. Ethnic identification of crania from measurements. Papers of the Peabody Museum of Archaeology and Ethnology, Vol. 82. Cambridge, MA: Harvard University. p 1–108.

Howells WW. 1996. Howells' craniometric data on the Internet. Am J Phys Anthropol 101:441–442.

Huberty CJ. 1994. Applied discriminant analysis. New York: Wiley.

Huberty CJ, Olejnik S. 2006. Applied MANOVA and discriminant analysis. New York: Wiley.

Hughes S, Wright R, Barry M. 2005. Virtual reconstruction and morphological analysis of the cranium of an ancient Egyptian mummy. Australas Phys Eng Sci Med 28:122–127.

Jantz RL, Ousley SD. 2005. Fordisc 3: computerized forensic discriminant functions. Version 3.0. Knoxville: The University of Tennessee.

Jorde LB, Wooding SP. 2004. Genetic variation, classification, and 'race'. Nat Genet Suppl 36:S28–S33.

Klepinger LL. 2006. Fundamentals of forensic anthropology. New York: Wiley.

Konigsberg LW, Jantz RL. 2002. Mixture analysis as an alternative to "determination of ancestry". Am J Phys Anthropol Suppl 34:96.

Konigsberg LW, Ousley SD. 1995. Multivariate quantitative genetics of anthropometric traits from the Boas data. Hum Biol 67:481–498.

Lewontin RC. 1972. The apportionment of human diversity. Evol Biol 6:381–398.

Lieberman L, Kirk RC. 2004. What should we teach about the concept of race? Anthropol Educ Q 35:137–145.

Livingstone F. 1962. On the non-existence of human races. Curr Anthropol 3:279–281.

Mahalanobis PC. 1936. On the generalized distance in statistics. Proc Natl Inst Sci India 2:49–55.

Maindonald J, Braun WJ. 2003. Data analysis and graphics using R—an example-based approach. New York: Cambridge University Press.

Ousley SD, Billeck WT. 2001. Assessing tribal identity in the Plains using nontraditional craniometrics (interlandmark distances). Am J Phys Anthropol Suppl 32:115–116.

Ousley SD, Billeck WT, Hollinger RE. 2005. Federal repatriation legislation and the role of physical anthropology in repatriation. Yrbk Phys Anthropol 48:2–32.

Ousley SD, Jantz RL. 1996. Fordisc 2.0: personal computer forensic discriminant functions. Knoxville: The University of Tennessee.

Ousley SD, Jantz RL. 1997. The Forensic Data Bank: documenting skeletal trends in the United States. In: Reichs K, editor. Forensic osteology, 2nd ed. Springfield, IL: C.C. Thomas. p 297–315.

Ousley SD, Jantz RL. 2002. Social races and human populations: why forensic anthropologists are good at identifying races. Am J Phys Anthropol Suppl 34:83–84.

Ousley SD, Jantz RL, Freid D. 2007. Exploring human craniometric variation: statistical, mensural, biological, and historical considerations. Am J Phys Anthropol Suppl 44:182.

Ousley SD, McKeown A. 2003. A comparison of morphometric data and methods in classification. Am J Phys Anthropol Suppl 36:162.

Ousley SD, Seebauer JL, Jones EB. 2003. Forensic anthropology, repatriation, and the "Mongoloid" problem. Proc Am Acad Forensic Sci 9:245–246.

Parra E, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD. 1998. Estimating African-American admixture proportions by use of population-specific alleles. Am J Hum Genet 63:1839–1851.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Relethford JH. 1994. Craniometric variation among modern human populations. Am J Phys Anthropol 95:53–62.

Relethford JH. 2002. Apportionment of global human genetic diversity based on craniometrics and skin color. Am J Phys Anthropol 118:393–398.

Roseman CC. 2004. Detecting interregionally diversifying natural selection on modern human cranial form by using matched molecular and morphometrics data. Proc Natl Acad Sci USA 101:12824–12829.

Roseman CC, Weaver TD. 2004. Multivariate apportionment of global human craniometric diversity. Am J Phys Anthropol 125:257–263.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. 2002. The genetic structure of human populations. Science 298:2381–2385.

Ross AH, Slice DE, Pachar DV. 2005. Forensic identifications and the complexity of determining biological affinities of "Hispanic" crania. Proc Am Acad Forensic Sci 11:292.

Ross A, Slice DE, Ubelaker DH, Falsetti A. 2004. Population affinities of 19th century Cuban crania: implications for identification criteria in Cuban Americans in South Florida. J Forensic Sci 49:11–16.

SAS Institute Inc. 2001. SAS/SHARE 9 user's guide. Cary, NC: SAS Institute Inc.

Sauer NJ. 1992. Forensic anthropology and the concept of race: if races don't exist, why are forensic anthropologists so good at identifying them? Soc Sci Med 34:107–111.

Slice DE, Ross AH. 2004. Population affinities of "Hispanic" crania: implications for forensic identification. Proc Am Acad Forensic Sci 10:280–281.

Spradley MK. 2006. Biological anthropological aspects of the African diaspora; geographic origins, secular trends, and plastic versus genetic influences utilizing craniometric data. Ph.D. dissertation. Knoxville: The University of Tennessee.

Spradley MK, Jantz RL, Robinson A, Peccerelli F. 2008a. Demographic change and forensic identification: problems in metric identification of Hispanic skeletons. J Forensic Sci 53: 21–28.

Spradley MK, Ousley SD, Jantz RL. 2008b. Evaluating cranial morphometric relationships using discriminant function analysis. Am J Phys Anthropol Suppl 46:199.

Systat Software Inc. 2004. Systat Version 11. Point Richmond, CA: Systat Software Inc.

Tabachnick BG, Fidell LS. 2007. Using multivariate statistics, 5th ed. Boston: Allyn and Bacon.

Ubelaker DH, Ross AH, Graver SM. 2002. Application of forensic discriminant functions to a Spanish cranial sample. Forensic Sci Commun 4:1–6. Available at:http://www.fbi.gov/hq/lab/fsc/backissu/july2002/ubelaker1.htm.

Vercoutter J. 1976. Mirgissa III, Les necropoles. Paris: Guenther.

Vidarsdottir US, O'Higgins P, Stringer CB. 2002. A geometric morphometric study of regional differences in the ontogeny of the modern human facial skeleton. J Anat 201:211–229.

Wescott DJ, Jantz RL. 2005. Assessing craniofacial secular change in American Whites and Blacks using geometric morphometry. In: Slice D, editor. Modern morphometrics in physical anthropology. New York: Kluwer Academic. p 231–246.

Williams FL, Belcher RL, Armelagos GJ. 2005. Forensic misclassification of ancient Nubian crania: implications for assumptions about human variation. Curr Anthropol 46:340–346.

Wright L Jr. 1995. Who's black, who's white, and who cares: reconceptualizing the United States's definition of race and racial classifications. Vanderbilt Law Rev 48:513–570.

Wright R. 1992. Correlation between cranial form and geography in Homo sapiens: CRANID—a computer program for forensic and other applications. Archaeol Oceania 27:128–134.