Phylogenetic Techniques and Methodological Lessons from Bioarchaeology

Gordon F. M. Rakita

Phylogeny and phylogenetic methods may be new to archaeology, but an interest in quantitative examination of cultural relatedness is not new in anthropology. The early twentieth century witnessed considerable interest in quantifiable representations of cultural similarity and difference—for example, Driver and Kroeber’s (1932) Quantitative Expression of Cultural Relationships—but the interest waned. The disenchantment with such studies no doubt was a result of a lack of integration within anthropology, itself perhaps an outgrowth of the increasing subfield specialization possible with the development of anthropological departments. Decreasing knowledge about developments in other subfields led to a decline in integrated research questions and research designs.

Over the past two decades, however, there has been increasing interest within anthropology in the development and application of explanatory models based on Darwinian evolutionary theory. This renewed attention has sparked a revitalized interest in quantitative assessments of cultural relatedness. As a result, phylogenetic studies have become more commonplace. My purpose here is to review phylogenetic approaches that have been used in physical anthropology. In particular, I highlight some of the insights that can be gleaned from many decades of phylogenetic analysis performed by bioarchaeologists—physical anthropologists who specialize in human skeletal remains from archaeological contexts. In this study, I focus on physical and morphological attributes; the following chapter (chapter 9) covers a discussion of genetic material derived from skeletal remains. I begin with a basic description of what biological-distance (biodistance) studies are and how they are conducted. I follow with a discussion of why such studies are relevant to cultural
 PHYLOGENETICS. I then provide a brief history of phylogenetic studies conducted on human skeletal materials from the American Southwest. This history provides examples of methodological hurdles that have been dealt with by bioarchaeologists. I end with a summary of some key technical issues that should be considered by anyone exploring phylogenetic methods of data analysis.

**Biodistance Studies**

Biodistance studies have been performed in one form or another for most of the twentieth century. The basic premise to biodistance analysis is the assumption that human physical similarity is an indication of shared ancestry. More specifically, bioarchaeologists assume that phenotypic similarity mirrors or is an indication of genotypic similarity. This similarity is assumed to be the result of gene flow between populations of interacting individuals. The sources of data for bioarchaeologists are the human skeletons excavated from archaeological sites. Buikstra et al. (1990:1) describe biodistance studies as examining "variation in bone or tooth shape and form in order to define patterns that are thought to reflect genetic relatedness within or between past populations." More colloquially, a person’s similarity to his or her parents and children is a direct reflection of the genetic material that they share.

Traditionally, two forms of data are used to quantify variation in bone and tooth morphology—data on metric traits and data on nonmetric traits (see Buikstra and Ubelaker 1994). Metric traits are commonly agreed upon, standardized linear measurements (or mathematical combinations thereof) of the distance from one point on the skeleton to another (e.g., Howells 1973). Examples of metric traits include the bizygomatic breadth—the distance between the most laterally situated points on the zygomatic arches—and the nasal index—the ratio of the width to height of the nasal aperture. Nonmetric traits, also referred to as discrete, epigenetic, or discontinuous traits, are variations in particular morphological features of the skeleton. Examples include the septal aperture—a small hole on the distal portion of the maxilla—and the mylohyoid bridge—a bone ridge that develops over the mylohyoid canal of the mandible.

Analysis of metric and nonmetric data derived from skeletal samples involves the statistical examination and evaluation of levels of similarity in the variation exhibited in those samples. Samples that exhibit widely divergent metric or nonmetric characteristics are assumed to represent individuals or populations that are widely divergent in their genetic makeup. That is, such divergence in phenotypic traits is considered indicative of limited gene flow between the populations represented by the samples. One specimen (a skeleton) can be compared to another specimen, sets of specimens (a skeletal series) can be compared to other sets, and individual specimens can be compared to multiple sets. For biodistance studies, at least three samples or sets are
needed so that the analyst can say that, for example, the biodistance between A and B is shorter than it is from either A or B to C. Biodistance is expressed either as a numerical estimate of similarity or dissimilarity and/or visually as spatial closeness on a phylogenetic tree or within two- or three-dimensional space.

As noted by Buikstra et al. (1990), biodistance studies can be conducted at a variety of analytical scales ranging from interracial, to interpopulation, to intraregional, to intrasite. Interracial studies attempt to assess the genetic distance between the “major races” of modern humans. Buikstra and colleagues found that such studies ceased to be published in the American Journal of Physical Anthropology after 1960. Interpopulational studies that examine biodistance among sets of skeletons representing national or state-level groups suffered a similar decline after 1960, although it was not as precipitous. Intraregional studies, on the other hand, increased dramatically after 1960 and represent a large proportion of biodistance studies published in the American Journal of Physical Anthropology. These studies concentrate on assessing microevolutionary processes using samples drawn from the same general geographic region. Intrasite studies, a growing but still minority approach in the 1970s and 1980s, represent analyses of the biological distance between specimens or sets of specimens excavated from a single archaeological site.

Several developments within the discipline impacted the nature of biodistance studies, perhaps the most pervasive of which was the establishment of new disciplinary goals for physical anthropology as a result of the influence of the evolutionary synthesis of the 1940s and 1950s. Washburn’s (1951) seminal article called for the establishment of a “new” physical anthropology that focused its attention on processes of evolution that have affected both extant and extinct primates (both human and nonhuman). This contrasted distinctly with the “old” physical anthropology, which was “primarily a technique” and was characterized by an “emphasis on classification based on types” (Washburn 1951: 298).

To be sure, physical anthropologists prior to the 1950s were well aware of the variation exhibited in the specimens that made up their “types,” but they did not see it as germane to their particular research questions. In terms of biodistance studies, the older period of research saw numerous descriptive studies of nonmetric traits, catalogs of summary statistics for metric variables for skeletal series, and a few examples of intraregional studies. Certainly intrasite studies were rare. Modern biodistance studies are far less concerned with classifying populations or specimens and more concerned with examining evolutionary processes affecting gene frequencies in samples. These processes include gene flow, genetic drift, population movements and migrations, and interpopulation interaction.

Related to the development of the new physical anthropology was the increasing distance anthropologists tried to establish between their research
and racist explanations or interpretations of their data (Brace 1982; Caspari 2003). Although biodistance relationships can be illustrated using phylogenetic trees, the use of such diagrams can suggest to the casual reader that speciation or a founding event has occurred. In many ways, the current lack of expressing biodistance relationships of modern *Homo sapiens sapiens* groups using such diagrams can be seen as a way of avoiding the appearance of typological thinking that harkens back to racist tendencies. Indeed, modern biodistance studies are often the target for recent claims that such research is inherently racist (Armélagos and Van Gerven 2003).

Finally, the development of novel multivariate statistical methods and the emergence of powerful personal computers hastened the advance of sophisticated quantitative applications. In his 1954 presentation of the coefficient of divergence, Spuhler (1954: 609) prefaced his discussion by stating that “it is not the statistical technique of preference. Discriminant analysis is the preferred technique for classificatory problems. . . . Unfortunately, discriminant methods involve both difficult and extremely laborious computations. When more than a few populations and measurements are used, the amount of labor with the usual desk computational equipment is prohibitive.” Things have changed to the point where most studies can benefit from powerful multivariate statistical procedures (e.g., Stojanowski 2003; Tomczak and Powell 2003), but there is a caveat: “Biological distance studies can become tedious mechanical exercises unless they are meshed within the context of evolutionary theory and historical problem solving” (Buikstra et al. 1990: 5).

### The Relevance of Biodistance Techniques

There are several reasons why bioarchaeological techniques and methodological advances in biodistance approaches are relevant to cultural phylogenetics research. First, both bioarchaeologists and archaeologists are interested in asking similar questions of their data sets. Both are interested in the interactions and movements of populations in prehistory. Whereas archaeologists have traditionally interpreted the magnitude of similarity in cultural traits as indicative of the magnitude of cultural interaction, bioarchaeologists have interpreted magnitude of phenotypic similarity as indicative of biological interaction. In effect, morphological similarity is assumed to represent genetic or cultural similarity. Thus one would assume that measures of similarity and the techniques of grouping similar samples or populations would be issues of mutual concern.

In some respects biodistance studies have outstripped the archaeological analysis of interactions among past human populations. As Konigsberg and Buikstra (1995: 203) note, much work has focused on the definition of archaeological boundaries from spatial distributions of artifact styles or the geographic organization of site types. . . . However,
boundaries can also be inferred through the study of past human skeletal features. In that the mechanisms for biological evolution are more readily understood than the processes defining culture change, human morphology can provide boundary definitions that are more easily interpreted than other classes of archaeological evidence.

Second, it has been suggested by those who reject evolutionary models for cultural phenomena that the reticulate nature of cultural transmission precludes a cladistic approach. Reticulations occur when the streams of cultural traditions merge. Kroeber in particular was concerned that reticulation made biological evolution a poor model for cultural phenomena. His perception of the differences between organic evolution and cultural evolution are illustrated in figure 8.1. However, reticulation also occurs in biological phenomena, as when genetic traits are passed between contemporaneous populations. Indeed, the possibility of genetic reticulation between *Homo sapiens sapiens* and *Homo sapiens neandertalensis* populations in Paleolithic Europe is still being debated (Klein 1999; O’Rourke 2003; O’Rourke et al. 2000; Wolpoff 1999). Bioarchaeologists are aware of these reticulations, and it is precisely these sorts of transmission processes that provide the basis for biodistance models.

Finally, both archaeologists and bioarchaeologists acknowledge that cultural transmission is often accompanied by or occurs together with genetic transmission. Thus the results of both cultural-distance and biodistance studies should be broadly comparable. Indeed, disjunctions between cultural and
genetic data should be illustrative of unique and potentially interesting situations. O'Brien and Lyman (2002: 27) have recently remarked that "the independence of biological and cultural evolution must be treated as a null hypothesis." Further, "it has long been observed that cultural transmission is independent of biological transmission, but this does not mean that cultural transmission will never be correlated with the degree of genetic relatedness between a transmitter and a receiver" (O'Brien and Lyman 2002: 27).

For these reasons, I would argue that an examination of bioarchaeological techniques and models for biological distance studies are extremely valuable to archaeologists interested in phylogenetic studies. Indeed, close cooperation between both sets of researchers promises to bear considerable fruit as each side learns from the other. In the meantime, what lessons have bioarchaeologists already learned that can usefully inform archaeological models of cultural phylogeny?

**Biodistance Studies in the American Southwest**

Beginning in the 1930s, bioarchaeological research in the Southwest, as well as in the rest of the country, began to undergo significant shifts in research agendas. Famous physical anthropologists such as Earnest Hooton, W. W. Howells, Aleš Hrdlička, Charles Seltzer, T. Dale Stewart, and James Spuhler tried their hands at quantitatively measuring the relatedness of various skeletal samples from the region. These researchers attempted to assess the relative genetic similarity of populations through examinations of observable metric features of the human skeleton, especially the cranium. Some of the crania from the ancestral Zuni site of Hawikku (referred to as "Old Zufii"), along with others from the Southwest, constituted the primary data sets for a series of craniometric studies published in the 1930s and 1940s (Brues 1946; Hrdlička 1931; Seltzer 1944; Stewart 1940; see also Hooton 1930). Several of these investigations were aimed at answering questions about the ancestry of the modern Zunis (a question originally raised by Frank Cushing). However, most also attempted to reconstruct biological affinities of the prehistoric and contemporary inhabitants of the greater Southwest.

Hrdlička (1931) made the first biodistance study in the Southwest, using cranial measurements from numerous locations, including southern Utah Basketmaker sites, Puyé in the Jemez mountains of New Mexico, "Old Zufii," Chaco Canyon, and Hopi Mesa. Using comparisons of various metric attributes of the crania, especially the cephalic index, Hrdlička arrived at several conclusions, including that the collections displayed two distinct morphological groups, one brachycephalic ("round-headed") and the other dolichocephalic ("long-headed"). Among the former fell the Utah Basketmaker specimens, and among the latter the Puyé and Hopi specimens. Additionally, Hrdlička noted that the geographic distribution of the two groups was unsystematic, which probably represented "considerable interpenetration."
Several years later, Carl Seltzer (1944) reanalyzed the collections, iteratively comparing the mean and standard deviation of over twenty metric traits and eleven indices of the skull for each pairing of the Hawikku collection against each other sample. He agreed with Hrdlička’s conclusion that the Zuni crania were morphologically similar to the Utah Basketmaker samples (Corruccini 1972). He further argued that the Zuni collection resembled those from Chaco Canyon. However, he also argued that

The supposedly sudden appearance of large numbers of deformed crania in the pre-Pueblo and the very earliest of Pueblo phases has caused the majority of archaeologists to believe that these deformed specimens marked the arrival of what they termed “a new race,” “a round-headed invasion.” The writer [however]... is prone to believe that the deformed crania are more the expression of a change in fashion or ideals of beauty rather than in physical type. (Seltzer 1944: 25)

Stewart’s (1940) analysis of skeletons excavated by Frank H. H. Roberts (1939, 1940) in the Zuni region lent support to this conclusion. Moreover, this interpretation of the skeletal remains challenged the traditional viewpoint of many archaeologists, including A. V. Kidder (1924), who suggested that the Basketmaker-Pueblo transition was marked by the arrival into the Southwest of a genetically dissimilar people.

Importantly, although Hrdlička and Seltzer (as well as Hooton and most other physical anthropologists of the day) referred to portions of their collections as belonging to this or that morphological “type,” this was not an exercise in mindless, essentialist classification. Nor should it be seen as a glimpse into suspected racist attitudes on the part of early twentieth-century physical anthropologists. No doubt such attitudes existed (Brace 1982), but Seltzer, for example, was quick to point out that terms such as “dolichoid” were descriptions of overall sample sets and often did not necessarily characterize the significant variation within groups: “The impression conveyed by these statements is that all Basket Maker crania are dolichocephalic, that is, have indices below 75, and that all the Pueblo crania are brachycephalic with indices over 80. This is not true” (Seltzer 1944: 26, emphasis added). Physical anthropologists on the whole were cognizant of the variation exhibited in their collections. It was simply that their research interests did not lead them to explanations of that variation.

One drawback to these early biodistance studies was the cumbersome and statistically invalid assessment of similarity in metric attributes. Seltzer’s iterative comparison of each of over thirty metric traits or indices from the Hawikku sample against over a half dozen other samples is illustrative of the mass of calculations necessary. Interpretation of the various results of such comparisons, together with concerns over the increasing experimental error rate, made individual researcher’s interpretations of results both subjective and suspect. In an attempt to alleviate these issues, and in response to growing
statistical sophistication, various multivariate measures of morphological similarity were developed. The most common early statistic used was the coefficient of racial likeness.

Although that coefficient was a computational improvement over previous methods, and although a few individuals continued to pursue craniometric biodistance studies into the 1950s, many of the techniques and assumptions of these research agendas were decried by both physical anthropologists (Stewart 1954) and archaeologists (Kraus 1954) alike. Stewart dealt a deathblow to the use of the coefficient of racial likeness and other similar measures in his critique of an analysis conducted by Spuhler (1954). Stewart (1954: 619) asked,

Is the coefficient of divergence (CD), the new statistical device which Spuhler uses, so much better than the CRL [coefficient of racial likeness]? By his own admission, it has about the same faults: Correlation of characters is ignored, characters are not weighted according to their importance, size of sample is not taken into account, etc. Why then does he proceed with this analysis, knowing as he surely does the nature of the criticisms which will be leveled at him? Part of the answer is, I believe, that he regards the means of analysis as secondary to the problem under investigation.

What was the result of such a critique? Did it drive the final nail into the coffin of biodistance studies? Actually, no. As Buikstra et al.'s (1990) study showed, biodistance studies remained a consistent, if low-frequency, topic from 1955 to 1985. Whereas after 1955 there was a decline in biodistance studies in comparison to studies of diet and disease, the later years of the study period showed an upswing. This renewed interest correlated with published studies regarding trait associations, sources of measurement error, the genetic basis of phenotypic traits, and discussions of distance statistics and models of gene flow related to migration and interaction (Buikstra 1980; Cheverud and Buikstra 1978, 1981a, 1981b, 1982; Cheverud et al. 1979; Konigsberg 1988, 1990; McGrath et al. 1984; Richtsmeier et al. 1984).

In the Southwest, later studies also examined the phenotypic similarity of a skeletal collection from one site with those from multiple other locations throughout the region—what Buikstra et al. (1990) termed intraregional studies. These included studies by Benfer (1968) and Butler (1971) of the Casas Grandes (Paquimé) material of northwestern Chihuahua; by Bennett (1973) of the Point of Pines burials; by El-Najjar (1974) on remains from Canyon de Chelly; by McWilliams (1974) on the Gran Quivira sample; by Heglar (1974) on Cochiti; and by Birkby (1973), Corruccini (1972), and Lumpkin (1976) on various samples. Such studies tailed off in the 1980s and 1990s, perhaps in response to growing political pressure surrounding the passage and implementation of the Native American Graves Protection and Repatriation Act. However, a new generation of bioarchaeologists is once again asking questions about the interaction (both culturally and biologically) and movement of prehistoric populations in the Southwest, and their reports are more fre-
Phylogenetic Techniques and Methodological Lessons

Methodological Lessons

Like most researchers studying phylogenetics, bioarchaeologists are careful to assess the selective pressures acting on possible traits used in biodistance studies. Traits whose frequency is heavily influenced by selection are more likely to yield information on biological adaptation rather than on population interaction. Biodistance studies, like analogous cultural-distance studies, require traits that are both heritable and influenced only by nonselective evolutionary forces such as drift. In this respect, the focus on neutral traits in archaeological analysis (Dunnell 1978; Hurt and Rakita 2001) may not be as misplaced as some believe. Stylistic traits, that is, those traits whose distribution in time and space is regulated by nonselective processes, are precisely the sorts of traits that would be most useful in cultural-distance models.

Of course, separating stylistic (neutral) from functional (adaptive) traits is not as simple as many of us would hope. As has been recently noted (Hurt and Rakita 2001; Hurt, Rakita, and Leonard 2001; Hurt et al. 2001), stylistic traits can become associated with functional traits such that they do not behave in a manner we would expect of traits not under selection. These sorted traits can have considerable impact on cultural-distance studies. For over thirty years, bioarchaeologists have noted the difficulties inherent in including traits that are correlated with each other. Indeed, it is standard in biodistance studies (e.g., Buikstra 1980; Konigsberg 2000; Pietrusewsky 2000) to check for intertrait correlations and to eliminate those traits that are found to be strongly associated with others.

Alternatively, proper selection of distance measures can eliminate or ameliorate the effects of trait correlations. Bioarchaeologists have turned to measures such as the Mahalanobis $D^2$ generalized-distance statistic in their examinations of the biological distance between populations. Generally speaking, Mahalanobis distance is the squared distance between two samples, taking into account any number of variables or characteristic states. One advantage of the Mahalanobis $D^2$ statistic is that it controls for intertrait correlations and mechanically satisfies the assumptions of a normal distribution of variables and an equivalence of covariance matrices. Mahalanobis distance has the added advantage of providing a single quantitative value that represents the dissimilarity between two groups, thus making it a useful first step in multivariate clustering techniques.

In a similar way, bioarchaeologists are careful to examine the level of correlation between traits used in their phylogenetic studies and other characteristics of their samples that may not be used in calculating biological distance. For example, traits highly correlated with the sex or age of individuals within their samples are eliminated. Again, the assumption is that such traits reflect...
developmental or sex-based etiology and not the genetic ancestry of an individual.

Bioarchaeologists are also quite wary about the possibilities of inter- and intraobserver error (Aftandilian 1995; Buikstra and Ubelaker 1994). These sorts of errors are introduced when two or more researchers collect information differently from each other or when one researcher collects information differently in different phases of the research. These sources of error have been largely neglected throughout archaeology’s history (but see Boyd 1987; Fish 1978; Whittaker et al. 1998).

The translation of genotypic information into phenotypic expression is also of concern to bioarchaeologists. For example, many nonmetric traits exhibit what is referred to as “quasi-continuous” expression, or traits that have a “threshold model” of expression (Konigsberg 2000; Saunders 1989). Such traits are controlled by continuously distributed genotypic attributes, but phenotypic expression does not occur until that genotypic attribute exceeds a threshold. Thus what in reality is a continuously distributed genotypic trait is phenotypically expressed dichotomously. As Konigsberg (2000) points out, it is the underlying genotype that is of interest to researchers, not necessarily the phenotypic expression. Yet it is the phenotypic expression that is available for study. In like fashion, archaeologists attempting phylogenetic reconstruction of prehistoric cultures should consider how cultural ideas are translated into cultural materials. Schiffer and others (e.g., Schiffer 1999; Schiffer and Skibo 1997; Schiffer et al. 2001) have been building the literature on such issues, which should be consulted.

Numerous other methodological issues can also impact archaeological phylogenetics. For example, most phylogenetic case studies involve the use of intricate multivariate techniques, yet rarely are the assumptions of such statistics discussed. Grouping methods abound, but what are the advantages and disadvantages of each? Why is one chosen over another? And what do we do about differences in sample sizes? Randomization techniques, which have become common in bioarchaeological circles, may be helpful with this issue and may also provide useful methods for dealing with missing data. Discussions of these issues are found in archaeology, for example in Neff’s (2002) discussion of quantitative techniques for analyzing ceramic compositional data. Bioarchaeologists have also struggled with these and other issues and have their own insights to add.

Conclusion

I would not argue in a similar vein as Stewart (1954) that archaeologists conducting phylogenetic studies regard phylogenetics as secondary to the problem under investigation. Quite to the contrary, phylogenetic methods are becoming more commonplace in archaeology. As archaeologists begin to use cladistics and other methods (e.g., O’Brien and Lyman 2003a), they can draw
important lessons from others who share their interest in the cultural and biological history of human populations. Bioarchaeologists have been grappling with how to reconstruct the interactions and movements of prehistoric populations for over a century. Archaeologists might do well to heed the methodological mistakes, hurdles, improvements, and developments within the biodistance studies of bioarchaeology.

Equally important, however, archaeologists should not lose sight of the research questions they are asking. Stewart’s concern was that bioarchaeologists in the 1950s were placing undue emphasis on the questions they were asking and not focusing enough upon the methods they were using. Conversely, I recommend that researchers using phylogenetic methods and techniques continually reassess their research goals in order to ensure that they are using the correct methods. Archaeologists should work to ensure that phylogenetic techniques are indeed answering the questions they wish to ask and are not simply becoming, as Buikstra et al. (1990: 5) put it, a “tedious mechanical exercise” devoid of theory.