Multivariate Approaches to Development and Evolution

Gunther J. Eble

SFI WORKING PAPER: 1999-05-035

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

©NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author’s copyright. These works may be reposted only with the explicit permission of the copyright holder.

www.santafe.edu
Multivariate approaches to development and evolution

Gunther J. Eble

National Museum of Natural History
Smithsonian Institution
MRC-121, Washington, D.C. 20560, U.S.A.

and

Santa Fe Institute
1399 Hyde Park Road, Santa Fe, NM 87501, U.S.A.

INTRODUCTION

Development is the usual arena of evolutionary change in morphology, and much of evolution obviously depends on it. It is very common, however, for discussions of human evolution and developmental change to focus on individual characters (brain size, body weight, bipedalism, etc.). This particularist approach has been enormously useful in understanding how individual phenotypic modules have changed in the course of human evolutionary history. But it has also led to dilemmas when from the pattern of developmental evolution of individual features broad statements about evolution in general are made (e.g., paedomorphosis vs. peramorphosis). Dissociated heterochrony (McKinney and McNamara 1991) and heterotopy (Zelditch and Fink 1996) are natural consequences of the modular nature of most organisms, and they have often been documented. The relative dissociability of phenotypes (quasi-independence of Lewontin 1978) justifies a more comprehensive treatment of developmental change whenever synthetic statements are sought. Conversely, high morphological integration may well entail uniform developmental change across characters. A multivariate perspective permits the discernment of such alternatives, and the production of average descriptions of development and evolution in multidimensional space. Two approaches are immediately obvious by contrast to the particularist approach: the tabulation of how developmental change occurs across individual characters (a useful line of attack which often figures in the literature, but which soon becomes intractable as the number of characters increase and as redundancy confounds the potential for synthetic inferences), and the direct summarization of multivariate variation with statistical techniques. This paper will address the latter. McKinney and McNamara (1991) and Tissot (1988) give good overviews of the advantages of multivariate analysis in the context of heterochrony, and Shea (1985) and Klingenberg (1996) discuss multivariate allometry. While I will elaborate on traditional multivariate methods in this paper, my intent is to eventually move beyond approaches which have now become de rigueur, such as
principal components analysis, and explore alternative ways of assessing patterns and processes of human developmental change in a multivariate framework.

Evolutionary anthropology as a discipline, both in biological and paleontological terms, shares much with evolutionary biology and paleobiology, and yet for a plethora of reasons (both sociological and historical) has remained self-contained and somewhat isolated from methodological and conceptual developments in those fields. The reverse is also true, but from the standpoint of anthropology this isolation has meant a rather idiosyncratic application of such developments. This is because they are only relevant to the extent that they can be fruitfully adapted to the empirical situations evolutionary anthropology is usually faced with, and tied to the tradition of empirical research in a straightforward fashion. Landmark-based approaches, for example, are finally gaining wide acceptance within anthropology (e.g., Lynch et al. 1996; Chaline et al. 1998). Still, a number of recent advances in multivariate biology and paleobiology have not yet been explicitly applied in evolutionary anthropology.

Most discussions of evolutionary change in development focus on the fundamental notion of ontogenetic trajectory -- how it is modified from ancestor to descendant. If one had access to a fully resolved phylogenetic tree, and if complete ontogenetic sequences were available for all species, one would proceed to interpret transition after transition in terms of how particular trajectories were modified. This tradition, while defensible (it will be applied below), hides the possibility that other observables might be informative in understanding the relationship between development and evolution. For example, developmental constraints are usually taken as a given and their potential breaking deemphasized, since heterochrony is usually thought to occur along an established trajectory. A focus on individual transitions also impedes a comprehensive appraisal of sweeping cross-phylogenetic predictions, such as Von Baer's laws. I will argue below that an ensemble approach, explicitly directed at identifying statistical regularities across a clade, may give insights that would not be immediately obvious to a student of individual
transitions. From the perspective of human evolution studies, this implies incorporation of principles and methods most commonly found in invertebrate evolutionary paleobiology. In particular, the concepts of morphospace and disparity (Raup 1966; Foote 1997) are implicit in much of the data and aims of human evolution studies -- their explicit recognition could add new dimensions of interpretability. By adapting the notion of morphospace and disparity to the arena of development (see Eble 1998a, 1999a), I intend to suggest how such concepts can be valuable in the study of human developmental change and, by extension, human evolution in general.

Heterochrony is a powerful mechanism of evolutionary change in development. However, as much as in a univariate context (Rice 1997), in a multivariate context the question of whether additional ways of studying developmental evolution are possible or useful deserves more attention. Ultimately, all changes in development must involve changes in rate or timing (whether uniform or non-uniform, along a single trajectory or leading to novelty, having a spatial component or not) -- the question is whether any given formalism should be used by default or whether a more pluralistic attitude is desirable. Current discontent with the standard nomenclature of heterochrony has been for the most part cast in terms of univariate situations, in line with the received typology (Alberch et al. 1979). Exceptions include Zelditch and Fink (1996) and Godfrey and Sutherland (1996). While multivariate heterochrony will inevitably surface in the discussions, this paper will mostly be concerned with the general utility of multivariate approaches to developmental change, on the one hand, and with the possibility of, through them, studying evolution and development in novel ways, on the other.

I will proceed by first discussing the general concepts of morphospace and disparity. I will then briefly review some techniques for the multivariate description of variation. The concepts of developmental morphospaces and developmental disparity will then be introduced. Finally, I will apply such concepts and some of the techniques in an

MORPHOSPACE AND DISPARITY

In paleobiology and macroevolution, the use of multivariate approaches has over the years become encapsulated in two related concepts: morphospace and disparity. Morphospace studies have a rather long history. Any representation of variation in state space (uni- or multivariate) is tantamount to a morphospace representation. While in many fields, including biological anthropology, this is already implicit in the very application of statistical methods to describe and summarize variation, in paleobiology and macroevolution the idea of morphospace came to symbolize a concern for the relationship between actual and possible, and the mechanisms underlying overall distributions in ordination space. Hence the distinction between theoretical (Raup 1966, 1967, 1968; McGhee 1998) and empirical (McGhee 1991; Foote 1993a, 1997) morphospaces.

Generally conceived, the idea of morphospace often demands a multivariate perspective, because the usual goal is the representation of whole organisms, or representative samples of overall morphology, in state space. As in multivariate statistics, each original variable is treated equivalently, so as to allow many variables to contribute to the characterization of objects in morphospace (Foote 1991a).

The concept of disparity, in turn, grew out of phenetics and the notion of distance in state space (Sneath and Sokal 1973; Van Valen 1974, 1978; Cherry et al. 1982; Foote 1991a). Rather than a summarization of morphological distributions through ordination, the goal is to produce state variables which represent the average spread and spacing of forms in morphospace. Many morphospace studies nowadays thus incorporate the quantification of disparity as a major goal (e.g., Foote 1991b; 1992a, 1993a; 1997; Roy 1994, 1996; Wagner 1995, 1997; Eble 1998b, 1999b). A body of theory regarding
expectations for the dynamics of disparity vis-à-vis diversity is also beginning to grow (Foote 1993a; 1996).

In the end morphospace and disparity studies, beyond the reliance on multivariate statistics, seek to understand the evolution of whole clades through time, with taxa treated as equivalent entities subject to quantitative representation. While this has usually implied sampling fossils over several geological intervals, the concepts of morphospace and disparity are equally applicable to neontological data and to situations where time is measured differently (say, with phylogenies). Because morphospace and disparity studies can generate much insight into the mechanisms of macroevolution (including selection, developmental constraints, and chance), the challenge of extending them to other disciplines seems a worthy one.

MULTIVARIATE DESCRIPTION OF VARIATION

As implied, the concepts of morphospace and disparity do not in themselves demand a multivariate perspective. A multivariate approach does not supersede univariate ones. It is best seen as a complement that allows reduction of redundancy, power of summarization, or both. In addition, certain outliers are never revealed in an univariate context because they are by their very nature multivariate (see Barnett and Lewis 1994). While there are disadvantages to multivariate analysis (see Pimentel 1979; McKinney and McNamara 1991), they are far outweighed by the synthetic insight that usually accrues. And because form is multidimensional, morphospaces and disparity are best studied with multivariate methods.

Ordination

Multivariate ordination techniques can be used to produce space diagrams of one’s data, thus effectively approximating a morphospace representation. While some distortion of actual similarities can occur, points that are close together can be taken as similar entities
and vice-versa. Many ordination techniques can isolate differences on different axes (e.g., size on one axis and shape in others). What is usually referred to as "traditional morphometrics" (the multivariate analysis of linear distance measurements) relies heavily on ordination to gain insight into data structures. Methods that rely on landmark data, such as thin-plate splines, also have ordination as one main goal of the analysis (see Bookstein 1991). Here I will focus on principal components analysis because it is a straightforward technique that is widely used and that is immediately applicable to both linear distance measurements (on which most studies in evolutionary anthropology are based) and landmark data. Rohlf and Bookstein (1990), Bookstein (1991) and Marcus et al. (1996) contain discussions of other ordination techniques, including those designed to landmark data. More traditional methods can be found in modern textbooks on multivariate statistics (e.g., Pimentel 1979; Morrisson 1990; Reyment and Jöreskog 1993).

Principal components analysis (PCA) is probably the most widely used ordination technique in multivariate biology and paleobiology. Its appeal derives from its effectiveness in reducing the dimensionality of the data while at the same time summarizing much of the variance in the few resultant dimensions. The elegance of the underlying logic, which is based on linear algebra, allows a variety of extensions, elaborations and applications. In this sense PCA is often the method of choice when one needs to organize a multivariate array of data and then choose one of several pathways of further inquiry.

PCA is an eigenvector technique, and as such it shares many of the features of a whole class of methods, including factor analysis, canonical analysis and correspondence analysis, which attempt to unravel simple patterns in a set of multivariate observations (Davis 1986). It relies on the assumption that the original data are correlated in some way, so that its variance-covariance or correlation structure can be used as the basis for a dimensionality reduction process that retains as much as possible of the original variation (Jolliffe 1986).
Algebraically, the main objective is to create a set of new and uncorrelated variables (the principal components) that can be ranked according to the amount of variation explained. The iterative process of finding new variables involves maximizing the original variance explained by each of them. Geometrically, what PCA does is to define the principal axes of an ellipsoid encompassing the whole cloud of points in multidimensional space. In conformity with the algebraic formulation, one creates a set of new and orthogonal axes (the principal component axes) which are successively oriented according to the largest variance among individuals or variables (Neff and Marcus 1980).

Operationally, the finding of new variables is achieved by calculation of eigenvalues and eigenvectors from a square symmetric matrix representing the data in some form. This is done by solution of the characteristic equation. Each eigenvalue represents a proportion of the total variance. The elements of each eigenvector (loadings) are the coefficients of the linear equation that lead to the creation of a new variable from the linear weighted compound of the original variables. The eigenvectors give the orientations and the eigenvalues give the magnitudes of the principal axes (Neff and Marcus 1980; Davis 1986). The original objects or variables can be projected onto each PC axis as scores, which represent the importance of each axis in the objects or variables. It must be noted that the new variables and axes are latent (Bookstein et al. 1985; Morrison 1990), in the sense that they constitute abstract constructs that themselves might represent, in potentia, possible underlying causes.

The success of a PCA is more than anything else proportional to the degree of reduction of dimensionality. In other words, when the first few axes account for most of the total variance (exactly how much is a matter of technical debate), we are justified in presuming that a parsimonious representation of the original data structure was achieved (Morrison 1990). Although PCA may lead to the formulation of hypotheses, it is best conceived as a mathematical, not a statistical method (Davis 1986). Its primordial role is to represent data in simple form; whatever possible causal relationships are indicated, they
should not arise as a result of the logic of the method. Biological interpretability should be
the ultimate guide in inferences from PCA (Tissot 1988).

Cluster analysis
Cluster analysis is one of the simplest multivariate techniques. This stems in part
from an intuitive underlying logic and an easily interpretable graphical output. Apart from
its weaknesses (sensitivity to choice of similarity measure and method of clustering) and
relative primitiveness, its popularity implies that more researchers (sometimes even in
different areas) can readily understand results from a cluster analysis. This means that
whenever there is a suspicion that pattern is present in a strong way, one can benefit from at
least preliminarily undertaking such an analysis.

Given a multivariate space, one may be interested in the possibility of its
segregation into groups and the way those groups can be compared and associated. Cluster
analysis allows such an organization of hyperspace by segregating entities or variables into
groups, or clusters, that imply greater phenetic resemblance within than among groups. In
other words, the elements of each particular group produced are more closely related (in a
phenetic sense) than elements of different groups (Sneath and Sokal 1973; Pimentel 1979;
Neff and Marcus 1980).

Usually, cluster techniques are sequential, agglomerative, hierarchical and
nonoverlapping (hence the acronym of SAHN techniques). Sequential, because there is an
iteration of certain algorithmic calculations; agglomerative, because elements are added in
such a way that progressively fewer and larger clusters are formed; hierarchical, because
nested groups are formed, so that for N elements N-1 clusters are formed; and
nonoverlapping, since elements at any particular level are mutually exclusive -- clusters at
the same level never share elements (Sneath and Sokal 1973; Pimentel 1979).

Simply put, a cluster analysis involves an assessment of distance of objects or
variables on the basis of their values in a data matrix, the creation of clusters by using a
particular cluster algorithm, and the generation of a scaled tree-like graphic output, that is, a phenogram. Once this is done, the two-dimensional output can be interpreted in light of external data. Details on different similarity measures and clustering algorithms can be found in Sneath and Sokal (1973).

**Disparity measures**

A number of disparity measures are possible, by reference to an extensive literature on taxonomic distances (e.g., Sneath and Sokal 1973). It must be noted that measures of multivariate distance are most useful, and intuitive, when the variables are equally weighted (Van Valen 1974), such that transformations of the data may be necessary. Because different disparity metrics are meant to reflect the same underlying morphospace, one can often find equivalence relations among different metrics (Van Valen 1974; Foote 1995, pers. comm.), with general patterns emerging regardless of metric (e.g., Foote 1995). Still, at its most basic level two fundamentally different aspects of morphospace occupation, and thus different disparity measures, should be recognized when estimates of dispersion are sought: the variance and the range. The variance captures the average dissimilarity among forms in morphospace; the range reflects the amount of morphospace occupied (Foote 1991a). In multivariate terms, these measures can be generalized as the total variance (the sum of the univariate variances) and the total range (the sum of the univariate ranges) (see Van Valen 1974, 1978). Although empirically the total variance and the total range tend to correlate (Foote 1992b), the relationship is not always proportional or monotonic, since the total range is very sensitive to sample size (Foote 1991a, 1993b). When sample size differs among samples under comparison, procedures like morphological rarefaction (Foote 1992b) should be used. Other multivariate statistics of variation exist (see Van Valen 1974, 1978; Foote 1990), but in general the total variance and the total range can effectively summarize dispersion in morphospace.
DEVELOPMENTAL MORPHOSPACES AND DEVELOPMENTAL DISPARITY

Following Eble (1998a,c, 1999a), I here expand the notions of morphospace and disparity to the realm of development. A developmental morphospace is any morphospace that reflects development by reason of either the variables or the entities depicted in such morphospace carrying meaningful developmental information (see also McGhee 1998). As such, the concept of developmental morphospace is a natural morphological extension of the notions of epigenetic landscape (Waddington 1957), epigenetic space (Goodwin 1963) and parameter space (Alberch 1989). It also appears explicitly in the work of Ellers (1993), although there the focus was on theoretical parameters, not measurements. Implicitly, it is also discussed by Rice (1998a,b) and Zelditch and Fink (1996). The explicit construction of developmental morphospaces is a necessary step in assessing the importance of development in structuring adult morphological distributions, and in identifying phenomena like heterochrony, heterotopy, and developmental constraints. Use of the term brings into focus the very aims of evolutionary developmental biology, and encourages the incorporation of ontogenetic data in multivariate studies of evolution.

Accordingly, developmental disparity can be formalized as the disparity among objects in any subregion of a developmental morphospace. Developmental disparity can be quantified either across taxa or within taxa in developmental time, depending on the purpose. Instead of changes in amount of variation in evolutionary time, the focus is shifted to changes in amount of variation in ontogenetic time.

At the limit, evolutionary and developmental renditions of morphospaces and disparity can be brought together, for one might be interested in tracking how changes in disparity and morphospace occupation through ontogeny themselves change in evolutionary time. While for sufficiently well-sampled paleontological time series this can determined directly against a time scale, in many cases the indirect route of mapping on a phylogeny can be used. An emphasis on developmental morphospaces and developmental
disparity, especially from a multivariate perspective, generalizes the playing field of evolutionary developmental biology by allowing, in addition to standard analysis of changes in rate and timing, the description of clade shape in ontogenetic time, the assessment of global developmental trends, the testing of broad predictions such as Von Baer's laws, and the description of potential links between ontogeny and phylogeny (Eble 1998c, 1999a).

AN EMPIRICAL ILLUSTRATION: REANALYZING HEINZ'S (1966) HOMINOID DATASET

Below I explore a number of methods and approaches previously discussed with an empirical illustration involving ontogenetic data for extant hominoids extracted from Nicole Heinz's (1966) study. The dataset, while incorporating a number of observations on fossil species and a discussion of post-cranial variation, is mostly concerned with ontogenetic variation in Recent skulls of the superfamily Hominoidea (Homo, Pan, Gorilla, Pongo, and Hylobates). Ontogenetic variation is partitioned into dental age classes (no teeth erupted; deciduous or milk teeth; first permanent molar erupted or M1; second permanent molar erupted or M2; third permanent molar erupted or M3). In the dataset, only the human sample yields measurements for the no teeth erupted class. The dataset is explicitly restricted to the genus level, with intraspecific variation disregarded in the case of modern humans, and intrageneric variation disregarded for gibbons and the great apes. No malformations or pathological individuals are considered. Sexual dimorphism is also disregarded and reputed uninfluential in an analysis at the genus level (Heinz further argues that sexual differences play a role only rather late in development, after the onset of puberty, and that intergeneric differences are mostly laid out earlier). In any case, the present data analyses and interpretations are limited by the resolution of the data.

While Heinz's study was mostly concerned with levels of variability, the dataset suffers from a rather pronounced unevenness in sample sizes across species. Thus, only
average values will be utilized in the present analyses, even though variance and coefficient of variation estimates are reported in Heinz's study as well.

Thirty-five distance measurements are presented in the dataset, but only 29 of these are consistently reported across species. Six of these measurements were not reported for the M3 dental stage in Gorilla, Pongo and Hylobates, and were accordingly not considered. Thus, 23 variables were consistently represented. However, the number of objects is restricted to 21 (4 dental stages per species plus 1 stage with no teeth erupted in humans). Whenever the number of objects is smaller than the number of variables, procedures such as principal components analysis give only approximate scores because of computational complexities. For the present purposes, 4 additional variables were thus removed; such variables correspond to the set of "non-classical measurements" in Heinz's work. The final data matrix then consists of 21 objects and 19 variables.

In exploratory fashion, several "developmental morphospace" and "developmental disparity" analyses were performed: a principal components analysis on log variables, with size clearly expressed on PC1; a principal components analysis on row normalized log variables, emphasizing shape and substantially removing size effects; cluster analyses, to identify phenetic propinquity among adults and juveniles of different taxa; juvenile-adult comparisons of disparity; mapping of developmental disparity against a phylogeny; and comparisons of ontogenetic trajectories of disparity relative to a Hylobates juvenile standard.

**Principal components analysis on log variables**

A principal components analysis was run on the correlation matrix of log variables. The choice between variance-covariance and correlation matrices is not a trivial one. Even though the difference between the two is only a matter of standardization (making the variance-covariance matrix a correlation matrix), they imply different manifestations of variation, and therefore lead to PCA results which are not interchangeable (Jolliffe 1986;
Johnson and Wichern 1992). While the use of variance-covariance matrices has become standard in many multivariate growth studies, it also gives greater importance to variables with high variance. The scale of measurement can thus significantly affect the results, either because of incommensurate units of measurement or because of widely different variances, as was the case here. In such situations, PC's are less informative -- the first few will summarize little more than the relative sizes of variances. In addition, informal comparison of PCA results from different studies is more straightforward when correlation matrices are used (Jolliffe 1986).

The first three principal components were retained for further analysis, based on (1) inspection of natural breaks in the decay of the variance of ranked eigenvalues, (2) amount of variance explained (98.6%), and (3) identification of PC's with more than one variable displaying relatively large values (while the retention of only the first two PC's would achieve the goal of variance summarization -- 95% --, interpretation of the third PC is biologically meaningful, as will be seen below).

PC I (84.2% of the variance) has all variables loading highly and positively on it. PC I is thus a size-allometry axis, which while also summarizing shape, is the best overall summary of size variation. PC II (10.8%) has palate length, basion-prosthion length, maximal biparietal width, porion-bregma height, and minimum frontal width loading highly. PC III (3.6%) is mostly influenced by variation in interorbital width, and to a lesser extent by foramen magnum length.

The resulting ordinations (see Fig.1), using scores on the rotated principal components, can be seen as representations of size-shape developmental morphospaces. PC I is here plotted against PC II and PC III. The PC I-PC II ordination clearly separates the ontogenetic trajectories of Hylobates and Homo from those of Pongo, Gorilla and Pan, which overlap to a substantial degree. It is only at larger sizes that the overlap on PC II is reduced (note positions of Pongo M3 and Gorilla M2 and M3). Note, on the PCI-PC II ordination, the isolation of Homo at the no teeth stage. In contrast, the PC I-PC III
ordination approximates *Homo* at the no teeth stage to *Hylobates*, on the one hand, and *Homo* at the milk teeth through M3 stages to the more derived apes. There is also less overlap among the ontogenetic trajectories of *Pongo*, *Gorilla*, and *Pan*. *Pongo* and *Pan* at the milk stage, for example, are quite separate in PC III, in contrast to their proximity in PC II.

On the assumption that the timing of molar eruption in apes is equivalent, and by taking into account the timing of molar eruption in humans and our current understanding of phylogenetic relationships among extant hominoids (for overviews, see Futuyma 1998; Lewin 1998), one can advance some heterochronic inferences (see Fig. 2). On PC I, great apes and humans differ from *Hylobates* mostly by predisplacement, *Gorilla* from *Pongo* by predisplacement, *Pan* from *Gorilla* by postdisplacement, and *Homo* from *Pan* by a combination of predisplacement and hypermorphosis (and, to a limited extent, neoteny). On PC II, *Gorilla* differs from *Pongo* mostly by acceleration, and *Pan* differs from *Gorilla* by neoteny and postdisplacement. The difference between *Homo* and *Pan* is describable in terms of a combination of neoteny and postdisplacement. Finally, on PC III, great apes differ from gibbons mostly by postdisplacement. *Pongo* in addition differs through non-linearity of the underlying ontogenetic trajectory, in other words, a different growth function implying non-uniform change from ancestor to descendant applies (see Rice 1997). A change into a qualitatively different growth function thus would also characterize the difference between *Gorilla* and *Pongo*, although much of the difference is also attributable to predisplacement. *Pan* differs from *Gorilla* by neoteny, and *Homo* differs from *Pan* by neoteny and predisplacement.
Figure 1. Principal components ordinations based on the correlation matrix of log variables. Age runs from left to right. Successive points correspond to the milk teeth, M1, M2, and M3 stages. In *Homo*, the no teeth stage is also represented.
Figure 2. Principal components based on the correlation matrix of log variables, plotted against age. The no teeth stage in *Homo* is not shown. "L" stands for the milk teeth stage.
Principal components analysis on row normalized log variables

Row normalization was used to partially remove the effects of size. Row normalization renders the sum of squares of variates for each object equal to one; it retains the proportionality of variables within objects, and destroys differences in magnitude between objects (Reyment and Jöreskog 1993).

As with the analysis on non-normalized log variables, the first three principal components were retained for further analysis, once again based on (1) inspection of natural breaks in the decay of the variance of ranked eigenvalues, (2) amount of variance explained (89%), and (3) identification of PC’s with more than one variable displaying relatively large values (the first three PC’s collectively contained the highest loadings for all variables).

In contrast to the analysis of log variables, PC I here summarizes a considerably smaller proportion of the total variance (48%). Several of the loadings have low values, and several are negative. Thus, PC I in this case is hardly a size-allometry axis in the conventional sense, or a good proxy for overall size. The contribution of shape variation to it is more conspicuous. PC I has nasion-opisthion length, maximal biparietal width, minimum frontal width, basion-bregma height, porion-bregma height, nasion prosthion height, palate length, and nose height loading highly (notice that some of these are the variables that contributed importantly to variation on PC II in the analysis of log variables). PC II (25.6%) has bi-auricular width, palate width, orbital height, and interorbital width with important contributions. PC III (15.4%) appears to reflect mostly variation in basion-nasion length, and foramen magnum length and width.

Corresponding ordinations (again based on rotated axes) are thus best interpreted not as guides to multivariate allometry per se (although they do reflect allometric differences), but as developmental morphospaces relatively unconfounded by size. Fig. 3 shows PC I plotted against PC II and PC III. The PC I-PC II projection isolates the ontogenetic trajectory of *Homo* from the rest. The trajectories of *Hylobates*, *Pan*, and
Gorilla follow a similar gradient, with that of Pongo somewhat offset. In the PC I-PC III ordination, much as with the previous analysis Homo at the milk teeth stage is close to Hylobates. While in the PC I-PC II ordination juveniles at the milk teeth stage are rather disjunct, PC III approximates juveniles of Pongo and Homo to Gorilla and Pan.

In terms of the timing of molar eruption in apes and humans and extant hominoid phylogeny, the following qualitative heterochronic inferences can be made (see Fig.4). On PC I, the great apes (but not humans) differ from Hylobates by predisplacement, Gorilla from Pongo also by predisplacement (as well as neoteny by the M3 stage), Pan from Gorilla mostly by postdisplacement, and Homo from Pan by both postdisplacement and neoteny. On PC II, Pongo differs from Hylobates only at the M1 stage, perhaps suggesting non-uniform change; Gorilla differs from Pongo by predisplacement, and despite the substantial overlap in the trajectories, Pan is slightly neotenic relative to Gorilla. Homo is distinct from Pan through a combination of neoteny, postdisplacement, and hypermorphosis. On PC III, in turn, postdisplacement seems mostly responsible for the difference between great apes and gibbons, with Pongo in addition differing through non-uniform change in ontogeny. Accordingly, a change in growth function accounts for the difference between Gorilla and Pongo, with acceleration also being involved. The trajectories for Pan and Gorilla are almost indistinguishable. Finally, Homo differs from Pan through neoteny.

While these ordination results are constrained by the limitations of the dataset, they do suggest a considerable amount of dissociated heterochrony, not only in terms of different composite variables being subject to different heterochronic processes, but also because different heterochronic processes may contribute to the pattern of evolutionary change in the same composite variable. Also, no consistent heterochronic trend is apparent as one moves from more primitive to more derived taxa. Could large-scale regularities be discernible by evaluating patterns of variation globally, in average terms? The analyses below address this possibility.
Figure 3. Principal components ordinations based on the correlation matrix of row normalized log variables. Age runs from left to right. Successive points correspond to the milk teeth, M1, M2, and M3 stages. In Homo, the no teeth stage is also represented.
Figure 4. Principal components based on the correlation matrix of row normalized log variables, plotted against age. The no teeth stage in Homo is not shown. "L" stands for the milk teeth stage.
Cluster analyses

Cluster analysis can be used to compress information on phenetic relationships into two-dimensional diagrams. This has its disadvantages, for some distortion of relationships implied by ordinations can occur. Its main advantage is the provision of an average characterization of groupings in morphospace, effectively accomplishing multivariate sorting of the data. Historically, cluster analysis has usually served the purpose of taxonomic discrimination. It is rather uncommon, thus, for cluster analyses to be performed on samples of both juveniles and adults. Examples include Boyce (1964; see also Gould 1977) for hominoids (including fossils), and David and Laurin (1996) and Eble (1998c, 1999a), for echinoids.

Caution must be exercised in performing and interpreting cluster analyses, because of sensitivity to choice of clustering algorithm and distance metric. Though such choices can usually be retrospectively justified, a sensible approach is to base inferences on patterns that emerge from several different analyses. For convenience, I will use the Euclidean distance as the index of association and explore the results of 6 different algorithms (see Sneath and Sokal 1973): UPGMA, WPGMA, single-linkage, complete linkage, unweighted centroid, and weighted centroid clustering. Z-scores on row normalized log variables were used (thus the same data as in the second PCA analysis). Two sets of analysis were performed: (1) on all ontogenetic stages across taxa; and (2) on a subset comprising the milk teeth and the M3 representative for each taxon (thus maximizing the potential distinction between juveniles and adults).

When the total sample is considered, all methods of clustering consistently isolate the whole ontogenetic trajectory of Homo in a separate grouping, as implied by the PCA on the correlation matrix of row normalized log variables. Hylobates is also consistently separated into a cluster, as partially suggested by the ordinations. Pongo is generally (but not uniformly across analyses) isolated from Pan and Gorilla, with the milk teeth and M1 stages and M2 and M3 stages always conjunct, respectively. Younger stages of Pan tend
to cluster with younger stages of *Gorilla*, the same being true for later stages. Thus, cluster analyses of the total sample seem to accomplish a mixture of taxonomic and ontogenetic discrimination. For illustration purposes, the output of the UPGMA analysis is shown in Fig. 5.

When the subsample of milk teeth and M3 representatives is considered, a rather different picture is suggested (almost uniformly across cluster analyses). Results of the UPGMA analysis are illustrated in Fig 5. *Homo* is still consistently separated, but it always clusters with juveniles of *Pongo, Gorilla* and *Pan*. The adults of these taxa form a separate cluster, away from the larger cluster that also usually incorporates a separate grouping of *Hylobates* L and M3. The clustering of the great ape juveniles is consistent with Von Baer's laws, and the joining of *Homo* suggests overall paedomorphosis in human cranial evolution.

This difference between the analysis of the total sample and the analysis of extremes ultimately reflects the contrast between the structuring implied by the endpoints of ontogenetic trajectories and the fluidity (or not) of such trajectories in morphospace. It is clear from the above that both kinds of analyses can be informative. However, prudence is needed because the exclusion of ontogenetic stages can conceal the taxonomic signal of ontogenetic variation.

**Developmental disparity**

In terms of developmental disparity, several relevant questions arise: how the disparity among taxa changes across ontogenetic stages (in other words, how disparity changes as one proceeds from juveniles to adults), how the disparity of such pooled samples changes with the level of phylogenetic inclusiveness (or how the addition of particular taxa affects disparity), how disparate are ontogenetic endpoints and how such endpoint disparity changes in evolution, and how disparity accumulates through ontogeny relative to a standard of comparison. I will approach these questions below. Because of
Figure 5. UPGMA cluster analysis based on z-scores of row normalized log variables. In (a), all ontogenetic stages across taxa are included. In (b), only the milk teeth and the M3 stage are included. Notice the separation of Homo in (a) and its clustering with great ape juveniles in (b). See text for discussion.
the small sample sizes involved (which upon resampling yield comparatively large error bars), the results shown can for the most part be interpreted only qualitatively. They are meant to indicate possible representations (and interpretations) of data on developmental disparity.

**Disparity across stages**

Fig. 6 presents changes in hominoid disparity across ontogenetic stages. Z-scores on row normalized log variables were again used. Disparity was measured as the total variance of original variables; as the total variance across the first three PC's; and as the variance in each PC considered individually.

Changes in disparity (both as the total variance of the original variables and of the PC's) across stages are rather subdued. One major expectation borne out of evolutionary developmental biology is that Von Baer's second law should hold -- in other words, that disparity in earlier stages should be smaller than in later stages. This is generally the case, but it is interesting to note that the exclusion of *Homo* entails a slight reduction in disparity between M2 and M3.

In terms of the variance along individual PC's, however, disparity changes are much more irregular. When *Homo* is included, the variance along PC I changes in general conformance to Von Baer's second law, but the same does not apply to PC II, where the variance decays steadily, and to PC III, where it goes up and down. When *Homo* is excluded, a different pattern accrues but no consistency emerges. The variance of PC I increases up to M2 but then decreases, that of PC II initially increases but then stabilizes, and that of PC III goes up and down much in the same way as when *Homo* is included. Thus, while in general (and on average) Von Baer's second law would seem to hold in hominoid evolution, for sizeable portions of the phenotype (as summarized by PC II and PC III) it seems to be broken. Similar results are reported for echinoids by Eble (1998c, 1999a).
Figure 6. Developmental disparity among taxa across ontogenetic stages, in terms of z-scores of row normalized log variables, of the first three principal components based on the correlation matrix of row normalized log variables, and of individual principal components with *Homo* included and excluded, respectively.
Developmental disparity and phylogeny

The use of a phylogenetic framework was implicit in the interpretation of a number of results so far presented. In the absence of direct temporal information from the fossil record, the branching sequence implied by a phylogeny can be used as a framework for mapping of developmental disparity, and for the interpretation of its evolution.

Against the hominoid phylogeny presented in Fig. 7 (which was also used above) , the disparity (once again in terms of z-scores of row normalized log variables) at the milk teeth (L) and the M3 stages at different levels of phylogenetic inclusiveness is presented along nodes. At all nodes, the disparity of juveniles is smaller than that of near adults, in conformance with Von Baer's second law. At the level of *Hylobates*, however, there is a change in the proportionality of disparity differentials: while M3 disparity is essentially the same as that of the next node, that of the milk teeth stage is higher, suggesting a somewhat larger evolutionary jump early in development between *Hylobates* and the rest.

![Figure 7. Developmental disparity and phylogeny. Mapped along the nodes is the developmental disparity among taxa at the milk teeth (L) and M3 stages, reflecting different levels of phylogenetic inclusiveness. Along the terminal branches the developmental disparity within taxa (disparity of ontogenetic endpoints L and M3) is displayed. Disparity is quantified in terms of z-scores of row normalized log variables.](image-url)
The disparity of ontogenetic endpoints (L and M3 stages) is also presented along each branch in the phylogeny. Here a rather strong pattern emerges: the disparity of ontogenetic endpoints in *Homo* is much smaller than in the other hominoids. In fact, bootstrapped error bars for *Homo* never overlap with those of the other taxa. The increase in ontogenetic endpoint disparity from *Hylobates* to *Pongo* suggests global peramorphosis. Global paedomorphosis from *Gorilla* to *Pan* is also apparent, and very clear from *Pan* to *Homo*.

**Ontogenetic trajectories of disparity**

The quantification of developmental disparity within taxa can be accomplished with a focus on the disparity of endpoints, but one may be interested instead in how developmental disparity changes along the whole of the ontogenetic trajectory of a taxon, and how such trajectories of developmental disparity themselves compare from taxon to taxon. Using the milk teeth stage of *Hylobates* as a standard of comparison, the trajectories of disparity (based on z-scores of row normalized log variables) in each taxon can be represented relative to size or age. The age-disparity diagram (Fig.8) allows inferences regarding "overall" or global trends in ontogeny, without the complications of a representation of morphospatial relationships. David and Laurin (1996) apply similar diagrams to sea urchins, but using superimposition techniques to derive estimates of morphological distance.

As ontogeny proceeds, there is a progressive increase in disparity relative to the primitive state in all taxa. It is noticeable that the rate of change in humans is very subdued, however: there is more change within the trajectory of *Hylobates* itself than within *Homo*. *Pan* and *Pongo* overlap substantially in their trajectories of disparity. In terms of conventional heterochronic nomenclature, and treating disparity as a variable, Fig. 8 suggests a global peramorphic trend (by predisplacement and acceleration) from *Hylobates* to *Pongo* to *Gorilla*, and a global paedomorphic trend from *Gorilla* to *Pan* (by postdisplacement and neoteny) and to *Homo* (mostly by neoteny).
least, humans are on average neotenic. Of course, this representation of ontogenetic change is silent about individual characters and not incompatible with, say, peramorphosis also playing a role in human evolution.

![Ontogenetic trajectory of disparity](image)

Figure 8. Ontogenetic trajectory of disparity. Developmental disparity here reflects the within-taxon disparity of successive stages relative to a primitive standard of comparison, the milk teeth stage (L) of *Hylobates*.

**CONCLUSIONS**

The present paper addressed a number of approaches to the study of multivariate developmental change in human evolution. While several of them are already current in biological anthropology, it is hoped that some of the less conventional approaches and language used (which are nonetheless more and more familiar to quantitative paleobiologists) can add a new dimension to the interpretation of the relationship between development and evolution in the context of human morphological evolution.

Perhaps noticeable in this paper was the lack of emphasis on heterochrony per se at each and every step of the discussion. This is less an attempt at avoiding current ambiguity
concerning the definition and scope of heterochrony (Godfrey and Sutherland 1995a,b; 1996; Zelditch and Fink 1996; Rice 1997) -- so long as a given set of definitions and methods is provisionally accepted there is always room for important insights to be gained, however non-definitive --, than a desire to build on past work on human evolution and developmental change and advance (however exploratorily) concepts and methodological approaches that might eventually prove useful in biological anthropology.

It is hoped that the results here presented as illustration can be reevaluated with the application of some of the approaches outlined to other taxa, larger sample sizes, more variables, and better age resolution. Different data collection protocols (e.g., landmark-based) are likely to yield more precision and more insight as well. Beyond the various limitations of the dataset used, the results and their interpretation stand as suggestions for future research. Multivariate approaches in general and the language of morphospaces and disparity in particular deserve to be more fully integrated into the analytical arsenal available to studies of human evolution and development.

ACKNOWLEDGMENTS

I thank J. Clark, R. Potts, L. Profant and B. Senut for useful advice, comments and suggestions on topics relating to human evolution, and B. David, D. Erwin, M. Foote, D. Jablonski, D. Rasskin-Gutman and M. Zelditch for encouragement and feedback on topics relating to developmental disparity and developmental morphospaces. Research was supported by postdoctoral fellowships from the Smithsonian Institution and the Santa Fe Institute.
REFERENCES


