Evolutionary Anthropology 21:151–165 (2012)

ARTICLE

The Shape of Human Evolution: A Geometric Morphometrics Perspective

KAREN L. BAAB, KIERAN P. MCNULTY, AND F. JAMES ROHLF

Study of morphological form is fundamental to the discipline of paleoanthropology. The size and shape of our ancestors' anatomical features have long been the focus of research on hominin systematics, phylogeny, functional morphology, ontogeny, variation, and evolutionary change. Early physical anthropologists relied on both qualitative descriptions of anatomical shape and linear measurements to assess variation among hominins. The seminal works of W. W. Howells[1] and C. E. Oxnard[2] helped to bring multivariate techniques to the forefront of physical anthropology. Howells' intention was the objective delineation of components of shape, which could then fuel further analyses and interpretations, as well as clarification of the ways that growth influences interindividual and interpopulational differences in shape. He expressed concern that previous comparisons of individual measurements did not capture the overall shape of the skull, which is "expressed by the relations *between* measurements."[1:3] Similarly, Oxnard recognized that a multivariate approach to the study of complex shapes allows "for such perturbations (e.g., variation and covariation)...that are difficult to evaluate by eye and impossible to reveal by measurement and simple analysis alone."[2:6] While multivariate methods offered clear advantages over univariate or bivariate representations of shape, the analysis of traditional morphometric measures such as linear distances, angles, and ratios, has limitations when it comes to quantifying the complex geometry of some anatomical structures.

The advent of landmark-based geometric morphometric methods narrowed the gap between the subjects of biometric analyses and the quantitative abstractions used to represent them (Box 1). In contrast to traditional morphometric data, landmark data preserve both the dimensions of the object and the spatial relationships among these dimensions[3] (Box 2). An important advantage of this approach is the capacity to address a greater number and, in some cases, different types of research questions. Another useful feature of geometric morphometrics (GM) is the ability to visualize shape differences in the physical space of the organism, rather than interpreting extensive tables of numerical results. This enhances an investigator's ability to interpret results within an evolutionary, functional, or ontogenetic framework, and to present those results in a manner that is easily digested and evaluated by the scientific community. While landmark-based analyses are not a panacea for every paleoanthropological debate, they comprise a powerful set of tools for quantifying and testing aspects of shape variation and covariation.

Bookstein[4] provides thorough documentation of the theoretical and historical context of GM; reviews of the practical applications of GM are also available.[5] More recently, Slice[6] reviewed the development and use of GM within the broader subfield of physical anthropology, including the contributions of anthropologists to the development of GM. Therefore, the present review focuses primarily on the contributions of GM to studies of human evolution, particularly in the areas of characterizing taxonomic variation, allometry, and development, including modularity and integration. We begin by providing an overview of GM methods and the processes whereby raw data are made ready for statistical analysis. We have framed this discussion in general terms; mathematical details can be found in more technical treatments elsewhere in the literature. This is followed by a frank discussion of the strengths and weaknesses of GM, then a digest of a few important topics within paleoanthropology that have benefited from GM analyses. We finally address possible future directions for GM within the study of human evolution.

Karen Baab is an Assistant Professor in the Department of Anthropology and the Interdepartmental Doctoral Program in Anthropological Sciences at Stony Brook University. Her research focuses on the evolution of the skull in human and nonhuman primates and understanding the underlying forces responsible for inter- and intraspecific variation in cranial form. Email: Karen.Baab@stonybrook.edu.

Kieran McNulty is an Associate Professor in the University of Minnesota's Evolutionary Anthropology Laboratory. His research interests in the evolution and diversification of apes and humans are addressed using quantitative analyses to study hominoid cranial variation and through field research in the early Miocene fossil sites of western Kenya.

F. James Rohlf is a John S. Toll Professor in the Department of Ecology and Evolution at Stony Brook University and is interested in multivariate methods and applications in biology, especially in geometric morphometrics. He co-authored the text *Biometry*, articles on statistical methodology, and a number of software packages.

Key words: shape analysis; paleoanthropology; systematics; allometry; integration; modularity

DOI 10.1002/evan.21320
Published online in Wiley Online Library (wileyonlinelibrary.com).

GEOMETRIC MORPHOMETRICS BACKGROUND

Data collection: Landmarks and Semilandmarks

The foundation of GM analysis is the use of landmark data. A landmark is a precisely defined point on

Box 1. Glossary

Bending energy – a concept borrowed from mechanics; in the context of GM, it refers to the amount of energy required to bend a flat, infinite, and very thin metal plate upward or downward at points corresponding to the locations of landmarks in a reference configuration. The signs and magnitudes of the displacements correspond to the differences in the x and y (and possibly z) coordinates between the reference configuration and a target configuration.

Centroid size – the most commonly employed measure of size in geometric morphometrics, this is calculated as the square root of the sum of squared distances of each landmark in a landmark configuration to the configuration's centroid.

Geometric morphometrics – a set of methods used to acquire, process, and analyze (using multivariate statistics) landmarks or semilandmarks defined by a set of Cartesian coordinates; these methods preserve the geometric relations among the landmarks throughout analysis, allowing for visualization of shape, including mean shapes and differences among groups or individuals.

Landmarks - precisely defined points in two or three dimensions, defined by some rule; they are usually chosen to correspond to homologous points that can be found in all specimens.

Landmark configuration – a collection of landmarks recorded from a single biological specimen.

Procrustes distance – the standard measure of the difference between two landmark configurations after Procrustes superimposition; typically calculated as the square root of the sum of squared distances between all corresponding landmark pairs.

Procrustes superimposition – the process by which two or more landmark configurations are registered so that differences due to orientation, translation, and scale are removed; the registration is achieved by translating specimens to the origin, scaling all configurations to unit centroid size and performing rigid rotations so that the sum of the squared distances between corresponding pairs of landmarks is minimized.

Semilandmarks – a series of ordered points that are positioned along curves or surfaces; often used when there are an insufficient number of homologous landmarks to fully characterize shape in that region.

Shape – the geometric properties of a landmark configuration that are invariant to orientation, translation and scaling.

Shape space – the multidimensional space in which each unique shape (a landmark configuration) is represented by a single point; most analyses take place in tangent space, the Euclidean space that is tangent to the mean configuration following Procrustes superimposition.

a specimen, the position of which is recorded by a set of x-, y-, and (for 3-D data) z-coordinates. Whether or not individual landmarks are homologous can be debated, but they are generally chosen based on a criterion of homology of the underlying structures (Box 2). Many features of interest to paleoanthropologists, however, cannot be sufficiently characterized using only homologous points. This is particularly true for anatomical regions that have few or no easily identifiable discrete landmarks (for example, neurocrania and bone shafts). Semilandmarks, a series of ordered points, were developed as a means of quantifying such structures: homologous curves, contours, or surfaces between landmarks (Box 2).[7] In these cases, the structure, rather than individual semilandmarks, is considered homologous. Because the initial spacing of semilandmarks is arbitrary, sophisticated algorithms have been developed to respace ("slide") semilandmarks to maximize their correspondence across a sample. The advantage of this approach has been demonstrated in structures that exhibit localized bending but, unfortunately, software to perform this sliding step is not yet widely available for 3-D semilandmarks. A further complication is that, as reviewed by Perez,[8] there are two distinct ways to slide semilandmarks. While minimizing bending energy results in smoother, less complex shape changes, the minimization of Procrustes distances may be more useful for statistical comparisons and is more consistent with minimizing Procrustes distances for the overall superimposition.

Procrustes Superimposition

Raw coordinates from landmark data cannot themselves be meaningfully compared across multiple specimens, as they encode information not just about shape, but also about location, orientation, and scale (Fig. 1). Differences in location and orientation result from differences in the arbitrary starting position and alignment of each specimen during data collection. In the context of GM, scale equates to differences in isometric, but not allometric, size. The nearly universal method used in GM to extract shape information from raw coordinate data is called a generalized Procrustes analysis (GPA) (Fig. 1). Briefly, this method first translates specimen configurations to a common location by superimposing their centroids (geometric centers), then scales each configuration to unit centroid size (centroid size = 1). The final step in a GPA involves standardizing the orientations by rigidly rotating all configurations until corresponding landmarks across all specimens are as close together as possible.[9,10] If the data include sliding semilandmarks, then an initial GPA is performed, followed by the sliding step, then a second GPA.

Once a GPA has been executed, superimposed specimen configurations can each be represented as a single point in a high-dimensional shape space. Each unique shape corresponds to a specific position in this

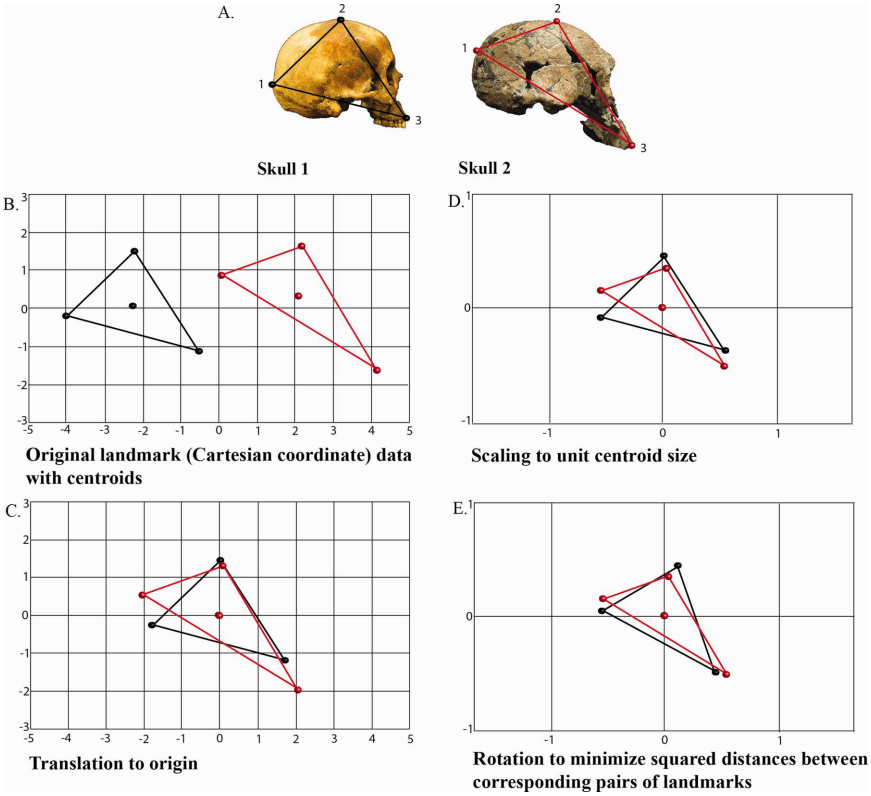

Figure 1. Procrustes superimposition. A. Three landmarks (1 = lambda, 2 = bregma, 3 = prosthion) are identified on two human skulls. B. For 2-D data, the landmark positions are recorded using two Cartesian coordinates (*x*,*y*). Coordinates of the centroid (geometric center) of each landmark configuration are calculated simply as the arithmetic mean of its *x*- and *y*-coordinates. C. Translation. By subtracting the centroid coordinates from the respective coordinates of each landmark in the configuration, differences in the location of each specimen are removed. This effectively moves the entire configuration so that its centroid is positioned at the origin of the coordinate system. D. Scaling. The centroid size for each specimen is calculated by first finding the distance of each landmark to its configuration's centroid, squaring these distances and summing them, and then taking the square root of the entire sum. Differences in scale are removed by dividing each coordinate of a configuration by its centroid size, thereby resizing each specimen to unit centroid size. While isometric size differences are now eliminated from the data, the original centroid sizes can be retained as a covariate to study size-correlated shape differences. E. Rotation. Differences in orientation are removed by rotating configurations around their centroids to minimize differences between each configuration and a reference (by minimizing the sum of squared distances across all corresponding landmarks). In an ordinary Procrustes analysis (N = 2 specimens; illustrated here), there is a direct solution whereby one specimen (Skull 1) is optimally rotated to fit the reference (Skull 2). In a generalized Procrustes analysis (N > 2 specimens), a mean configuration is first computed and all specimens are superimposed to this mean using an interative process. Coordinate data resulting from a Procrustes superimposition are now considered shape data, free of differences in location, scale, and orientation. (Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.)

space, and the cloud of points in shape space reflects the different shapes of the specimens in the sample: the more two shapes differ, the further apart they are in shape space.[4] The new coordinates of the superimposed specimens are free of differences due to location, orientation, and scale; however, the shape space occupied by these superimposed landmark configurations is non-Euclidean, violating a fundamental assumption of standard parametric statistical methods. Hence, to use such methods, the landmark configurations must be projected into a space where Euclidean geometry applies. This space is typically constructed tangential to the GPA space at the point of the mean shape, and is eponymously called tangent space (see Box 3 for more details regarding this projection).[11,12] An analogous projection that will be familiar to most readers is the mapping of the curved geography of the earth onto a flat 2-D surface.

Coordinates subjected to superimposition and projection into tangent space can now be analyzed with the usual multivariate statistical tools. These transformed coordinates are typically called shape variables (or tangent space coordinates). Thin-plate splines can also be used to visualize the tangent space (Box 4).

Procrustes Distance

The usual measure of the difference in shape between two superimposed landmark configurations is called a Procrustes distance. This distance is the square root of the sum of all the squared distances between corresponding landmarks in the two superimposed objects. Procrustes distance is the quantity minimized during GPA superimposition. It can also be interpreted as the distance between the two points representing the individual configurations in a multidimensional shape space (technically Kendall's shape space) (Box 3). Thus, Procrustes distance is a measure of differences in shape: a larger distance implies a greater magnitude of shape difference, while a distance of zero indicates that the two shapes are identical.

The use of the term "Procrustes distance," however, is somewhat imprecise in the literature. As described in Box 3, there are three shape spaces commonly associated with GM analyses: Kendall's shape space, named after the mathematician David Kendall, who outlined much of the theoretical underpinning of shape analysis,[13] is the space of all possible pairwise superimpositions; GPA space is the result of superimposing more than two specimens on the sample mean; and tangent space is the Euclidean space into which configurations are typically projected for analysis. The term "Procrustes distance" is only properly applied to the distance between configurations in Kendall's shape space.[12,14] However, most shape analyses occur in either GPA or tan-

Box 2. Landmarks and Geometry

Descriptive morphological analysis allows for the exhaustive characterization of a specimen's anatomy, which is an important component of any morphological study. However, qualitative descriptions do not provide objective criteria by which specimens can be compared to known ranges of variation and allow only coarse characterizations of variation and covariation among multiple taxa. For these reasons, many researchers turn to quantitative morphometric analyses that use rigorous epistemological criteria imposed by statistical analysis.

Univariate analyses provide entrée to the statistical toolkit, but capture only individual dimensions of the objects being studied. Multivariate analyses of linear measurements capture more of the objects' shapes, but do not retain the geometric relationships between measurements. Consider, for example, the simple interpoint measurements illustrated here (Fig. 2). Dimensions of basion-bregma and inion-glabella are easily obtained using calipers, and provide quantitative output describing two features of the specimen. The resulting distances, however, have lost their geometric relationship to each other (Table 1). The quantities "9.7" and "13.1" differ only in their magnitude and are uninformative about the angle between these chords, the point at which they intersect, and the distances between their endpoints.

The strength of landmark-based geometric morphometric analyses is that the original geometries of the specimens under study are preserved in the coordinate data. Rather than measuring the distance between pairs of points, geometric morphometric analyses use the landmarks themselves measured within a 2-D (x, y) or 3-D (x, y, z) coordinate system.

Going back to the same example, we could instead analyze the x-, y-coordinate data for the endpoints of the two linear measurements. Unlike the linear outputs, the coordinate data retain information on exactly how each of these landmarks relates to the others. Whereas the full geometry of the configuration is encoded in this set of four 2-D landmarks, six interlandmark distances would be necessary to capture this same information. Moreover, the discrepancy between the number of landmarks and the possible interlandmark distances increases as the number of landmarks increases. By retaining more of the biological information from the organism (that is, the geometric relationships among landmarks), more complex and subtle morphological questions can be addressed. Moreover, the preservation of geometric information within the data means that important morphological issues such as shape differences, vectors of variation, and axes of covariation can be easily visualized. And, of course, landmark coordinate data can be used to calculate all of the traditional continuous variables that were themselves defined by the landmarks: inter-point distances, angles, areas, volumes, and so forth.

The traditional "classes" of landmarks can also be seen in Figure 2. Bregma is a good example of a Type I landmark, where multiple

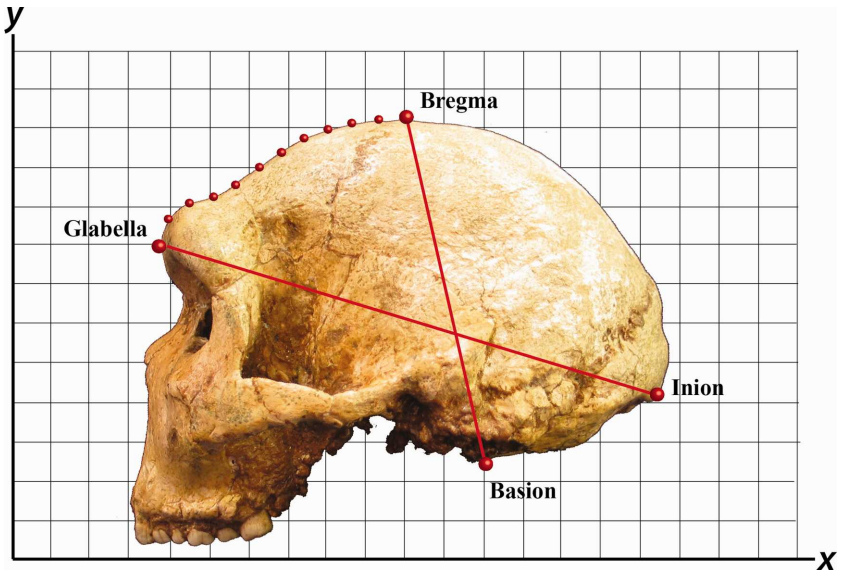

Figure 2. Example of two simple interpoint measurements, basion to bregma and inion to glabella. (Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.)

TABLE 1. Comparison of Measurement Data Derived from Linear and Coordinate Data

Type of data	Measurement	Data
Linear	basion - bregma	9.7
	inion - glabella	13.1
Coordinate (x, y)	basion	12, 2.4
	bregma	10, 11.3
	inion	16.5, 4.1
	glabella	3.8, 8.0

discrete tissues (the frontal and two parietal bones) intersect at a single point; the case for homology is strongest for Type I landmarks. Basion, on the other hand, is a Type II landmark; its homology is supported only by the geometry of the surrounding anatomy, the region immediately surrounding the foramen magnum. Glabella can be considered a Type III landmark in that its position, the most anterior midline point on the frontal bone when a skull is in the Frankfurt Horizontal position, is defined in part by distant anatomical structures (poria and orbitale).

Semilandmarks are illustrated in Figure 2 by the smaller spheres arrayed along the median plane between glabella and bregma. Unlike landmarks, which are individually presumed to be homologous across all specimens in a sample, homology in semilandmarks is presumed at the level of the structure or surface (here, the median plane of the frontal bone). This makes semilandmarks similar to Type III landmarks in that each is defined in part by the position of the semilandmarks on either side of it rather than by local anatomy. Thus, a semilandmark is not free to vary in any direction because it is bound in a series of other semilandmarks. This has implications for the overall degrees of freedom one has in statistical analysis.

gent space, and analogous measures in these spaces are usually labeled Procrustes distances as well. This imprecision in terminology is justified on the grounds that the distance between two configurations in any of the three shape spaces will be very similar, provided that overall shape variation is relatively small, as is the case in most biological analyses.[15]

APPLICATION OF GM METHODS

While the "morphometric revolution"[3] began more than 20 years ago, geometric morphometrics is still relatively new compared to traditional approaches, and research topics that are well suited for these methods may not always be clear to new practitioners. Broadly speaking, GM methods operate in the same realms as do other morphometric approaches, and can be used in both exploratory and hypothesis-testing frameworks. For example, a study may investigate shape differences between two extinct hominin species (exploratory), then go on to evaluate the hypothesis that a particular fossil should be assigned to one of these two species (hypothesis testing). The aims of most GM analyses fall into three main categories: characterizing and quantifying the main directions of shape variation or covariation in a sample; testing whether two or more groups differ significantly in some aspect of shape; or establishing the nature of relationships between shape and one or more additional variables.

In paleoanthropology, GM has been most frequently applied to questions regarding systematics, comparative anatomy, ontogenetic development, ecomorphology, and morphological integration. Theoretically, landmark-based morphometrics could also play a role in studies of biomechanics or phylogenetics, but historically they have had minimal impact in these fields.

Landmark data are particularly useful for recovering complex or subtle shape differences, which often characterize members of the same or closely related species or subsequent ontogenetic stages of development. Landmark data also provide opportunities for shape visualization that are not available using traditional morphometrics. However, there will be instances in which the effort required to collect, process, and analyze a large amount of landmark data outweighs the benefits of this approach. If the research question can be easily answered by collecting a small number of caliper-based measurements, then this tack will be both faster and computationally simpler. In fact, while subtle shape differences may be more easily identified using GM among very similar specimens, they may likewise be overwhelmed when comparing objects that differ greatly in other respects. Those subtle differences will still be encoded in the data, but it may take more sophisticated analyses to elucidate them.

Another disadvantage of GM analysis is its steep learning curve. Some analyses, such as principal components analysis, can be performed relatively easily with freely available and user-friendly software such as Morphologika, MorphoJ, or the Tps-series of programs, but more sophisticated analyses require more advanced geometric, algebraic, and statistical knowledge. There is also a fair amount of jargon associated with GM, which, in some cases, obfuscates what would otherwise be clear. For example, the axes associated with a two-block partial least squares analysis, an approach used in integration or modularity studies, are commonly referred to in non-GM contexts as "factors" or "dimensions." However, the term "singular warps" was coined for the situation where the two blocks of variables being analyzed are coordinate data.[17]

An additional strength of GM is the ability to mathematically separate shape variation from isometric size variation. Information about the original centroid sizes of landmark configurations is sequestered during the superimposition process, and can then be used to examine the relationship between size and shape. Isometric size differences are removed during superimposition, but allometric shape variation is unaffected. Therefore, studies that are explicitly interested in the relationship between size and shape may benefit from the use of GM methods.

The ability to collect data in the regions between homologous landmarks (that is, through the use of semilandmarks) is an important contribution of the GM toolkit, and opens many new avenues of investi-

gation. However, the use of semi-landmarks often results in a large number of variables, in some cases exceeding the number of individual specimens, which is problematic for standard statistical tests. A common "work around" is to subject the landmark data to a PCA, then use PC scores along a smaller number of PC axes as the variables for analysis. While this does reduce the number of variables being analyzed, neither biology nor statistics offers a clear rule-of-thumb for deciding how many PCs to analyze, and researchers use different criteria in making this decision.

Researchers embarking on a research project using GM data have several data collection techniques available to them. 2-D landmark data can be collected from photographs or x-ray images using freely available software such as tpsDig. However, many paleoanthropological subjects are ill-suited for 2-D analyses, since they vary in three dimensions and 2-D representations of 3-D objects can obscure anatomical relationships. For example, 2-D analysis of a gorilla cranium in frontal view would treat the nasal aperture as very close to the malar root because the length of the rostrum is obscured in that perspective. 3-D landmark data can be recorded directly from specimens using the portable Microscribe digitizer, a multi-joint mechanical arm with a stylus at one end. Alternatively, landmark data can be recorded from "virtual surfaces" using specialized software such as Landmark Editor or AVIZO. These surfaces might be rendered from CT or MRI data, or generated using laser or white-light surface scanners. In all cases, the goal is to generate a landmark configuration for each specimen that follows a common order and number of landmarks for the sample.

Systematics

Affinities of individual fossils

Since hominin taxa are generally defined by a combination of discrete and continuous characters, many of which relate to the shape of anatomical structures, morphometric data are frequently analyzed to address taxonomic questions. There is a long history of assigning individual fossils to taxa based on phenetic affinities established by traditional morphometric comparisons. The landmark-based assessments have shared similar success in this arena, as a few key examples will attest.

The cranium of the Late Miocene putative hominin *Sahelanthropus* (TM 266) has been the subject of both *in silico* reconstruction and GM analysis.[18,19] The initial publication of the cranial reconstruction[18] included some basic landmark-based quantitative analyses, which concluded that TM 266 belongs within the hominin lineage and that it may have used bipedal postures. A more extensive study by the same team[19] presented linear measurements as well as multivariate ordinations and clusters based on landmark data to support these initial conclusions. More recently in the hominin fossil record, the Late Pleistocene Hofmyer skull from South Africa, dated to 36 ka, was demonstrated to more closely resemble other contemporaneous humans from Europe than younger but geographically closer Khoe-San crania. This was done using a combination of PCA, canonical variates analysis, and UPGMA clustering.[20] The affinities of many other hominin specimens, including *Kenyanthropus*,[24] Indonesian *H. erectus*,[21] and early African *Homo*,[22] have been analyzed using similar tools.

The aims of most GM analyses fall into three main categories: characterizing and quantifying the main directions of shape variation or covariation in a sample; testing whether two or more groups differ significantly in some aspect of shape; or establishing the nature of relationships between shape and one or more additional variables.

GM is well suited for incorporating biological information about different morphogenetic factors such as ontogeny, size, and sex, when testing fossil affinities. These more sophisticated studies are becoming common as the methods mature within paleoanthropology. For example, the relationship between adult size and shape (static allometry) has been an important theme in GM analyses of the recently recovered fossil "hobbits" from Flores. The most extensive study to date was carried out by Baab and McNulty,[23] who used a series of phenetic, allometric, and asymmetry analyses to test hypotheses about the affinities of the LB1 cranium. Basic multivariate ordinations and clustering suggest a strong affinity between LB1 and crania of Pleistocene hominin taxa, including *H. erectus* and *H. habilis*. When scale was taken into consideration, the LB1 cranium was shown to conform to the pattern of morphology expected for a Pleistocene hominin, but not a modern human, of the same diminutive size. The same study also assessed asymmetry in LB1 by computing Procrustes distances between specimen configurations and their mirror images, rejecting the hypothesis that LB1's asymmetry is outside of the range of normal asymmetry.[23]

McNulty, Frost, and Strait[24] took a unique approach to assessing the taxonomic affinities of the Taung child by applying developmental trajectories from different extant taxa (apes and humans) to the juvenile specimen to estimate its adult morphology. Their results suggested that with regard to cranial shape, the estimated adult Taung specimen was most similar to the Sterkfontein *Australopithecus africanus*, and specifically to Sts 71 rather than Sts 5. They also reconciled some earlier contentious

Box 3. Spaces and Distances

The relationships among shapes comprising three landmarks in (A) Kendall's shape space, (B) GPA space ("Slice's shape space"), and (C) tangent space, are depicted in Figure 3. Landmark configurations lose degrees of freedom as a result of holding constant translation, rotation, and scaling during Procrustes alignment. Thus, any configuration of three landmarks can be represented as a single point in a 2-D space. Kendall's shape space can be visualized as the surface of a sphere and GPA space as the surface of a hemisphere; for configurations with more than three landmarks, these are, respectively, higher-dimensional spaces on the "surfaces" of a hypersphere and hyperhemisphere. Similarly, for configurations of three landmarks, Euclidean tangent space is 2-D space (that is, a plane), but becomes a higher-dimensional space for analyses with more than three landmarks.

The large black dot at the "north pole" represents the mean configuration from a GPA alignment. This shape is also used as the point of tangency for constructing the tangent space and, for convenience, is shown here as the reference for Kendall's shape space. The three white dots all represent the positions of a second shape in the different shape spaces. The black lines connecting these three points represent the projections of this point from Kendall's shape space to the GPA shape space to Euclidean tangent space. The projection to tangent space shown here is an orthogonal (perpendicular) projection, which best preserves the distances among specimens in Kendall's shape space.

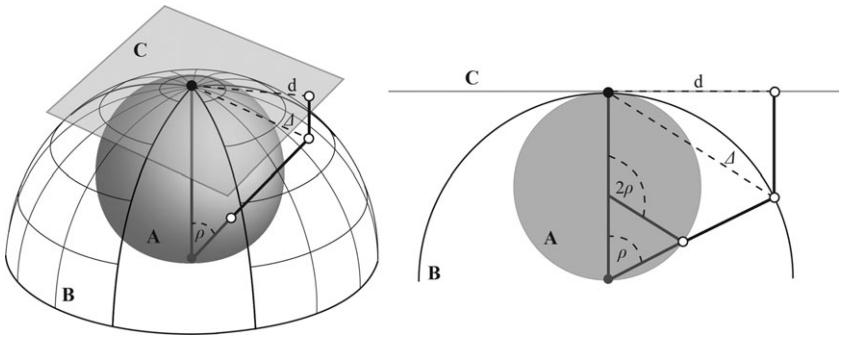

Figure 3. The relationships among shapes comprising three landmarks in (A) Kendall's shape space, (B) GPA space (``Slice's shape space''), and (C) tangent space.

If we were to section the figure on the left through the plane defined by the configurations' radii within GPA space (forming the angle ρ), we would see the simplified depiction to the right. As before, (A) represents Kendall's shape space, (B) is GPA space, and (C) is tangent space. Procrustes distance (ρ) is properly defined as the angular distance between two specimens in Kendall's shape space, the great-circle distance between two points on the (hyper) surface of Kendall's (hyper) sphere. Because specimens that lie in Kendall's shape space are scaled to a centroid size of 1, the great-circle distance between two points is equal to the angle (ρ, in radians) depicted here. Note that for these two configurations, the "south pole" of Kendall's shape space is coincident with the "center" of GPA space (gray dot).

The situation is slightly different for GPA-aligned specimens. Unlike specimens in Kendall's shape space, specimens aligned by GPA are optimally superimposed on the mean configuration. Thus, the only true Procrustes distances in GPA space are between a specimen and the mean. Distances between any other two specimens do not represent optimal alignment of those two; one can imagine that differences between two specimens could be reduced if they were aligned with only each other rather than with the sample mean. As in Kendall's shape space, when centroid sizes are scaled to 1 during GPA, the great-circle distance in GPA shape space is also equal to the angle (in radians) between the two points.

An alternate distance used by many is the value Δ, seen in the second diagram, which represents the straight-line (chord) distance between two points in shape space. While this is not a "true" Procrustes distance as such, it very closely approximates the great-circle distance, particularly when distances between specimens are small (as is nearly always the case in biological samples). Dryden and Mardia usefully distinguish these two metrics as Procrustes angular distance (ρ) and Procrustes chord distance (Δ). Of course, when the specimens are projected into tangent space, the "shape distance" between them is a simple Euclidean distance (d).

results from both linear and landmark studies of African ape ontogeny, showing that significant variation in some stages of development is not necessarily coincident with substantial differences in adult morphology.

Testing the single-species hypothesis in paleoanthropology

The degree and pattern of variation within single species can also be used as a benchmark for testing the null hypothesis that a sample of fossils represents a single species. If variation in the fossil sample is excessive relative to the comparator species, this can be interpreted as support for the alternative hypothesis, that the sample includes members of more than one species; like-

Box 4. Thin-Plate Splines

One of the most compelling contributions from GM is the TPS representation of shape differences between two specimens (Fig. 4). As Bookstein[16] noted, much of the morphometric work of the twentieth century has been implicitly trying to produce mathematically grounded methods with the explanatory power of D'Arcy Thompson's transformation grids. The TPS does this by providing an intuitive visualization that allows one to represent shape differences as a deformation of one shape into another, rather than as vectors or lists of numerical output. Unlike other methods for visualizing differences between sets of landmarks, TPSs provide continuous functions that interpolate shape differences in the regions between landmarks, much as a regression equation is a continuous function that can be used to interpolate the value of the y within the range of values of x. This makes it possible to visualize shape differences as deformations rather than separate displacements at each landmark. While this is a powerful visual tool, the shape changes in the interpolated regions cannot be interpreted literally because they are determined by the landmarks themselves.

As compelling as TPS grids are for displaying 2-D data, they are difficult to employ for 3-D visualization. Workers have tried three methods for depicting 3-D splines (as parallel stacks of 2-D grids, as an individual 2-D grid sectioned from a region of interest, or as an animation that moves a 2-D grid through a third axis, showing concomitant grid changes along its course), but none of these approaches has yielded the same impact and explanatory power of 2-D TPS grids.

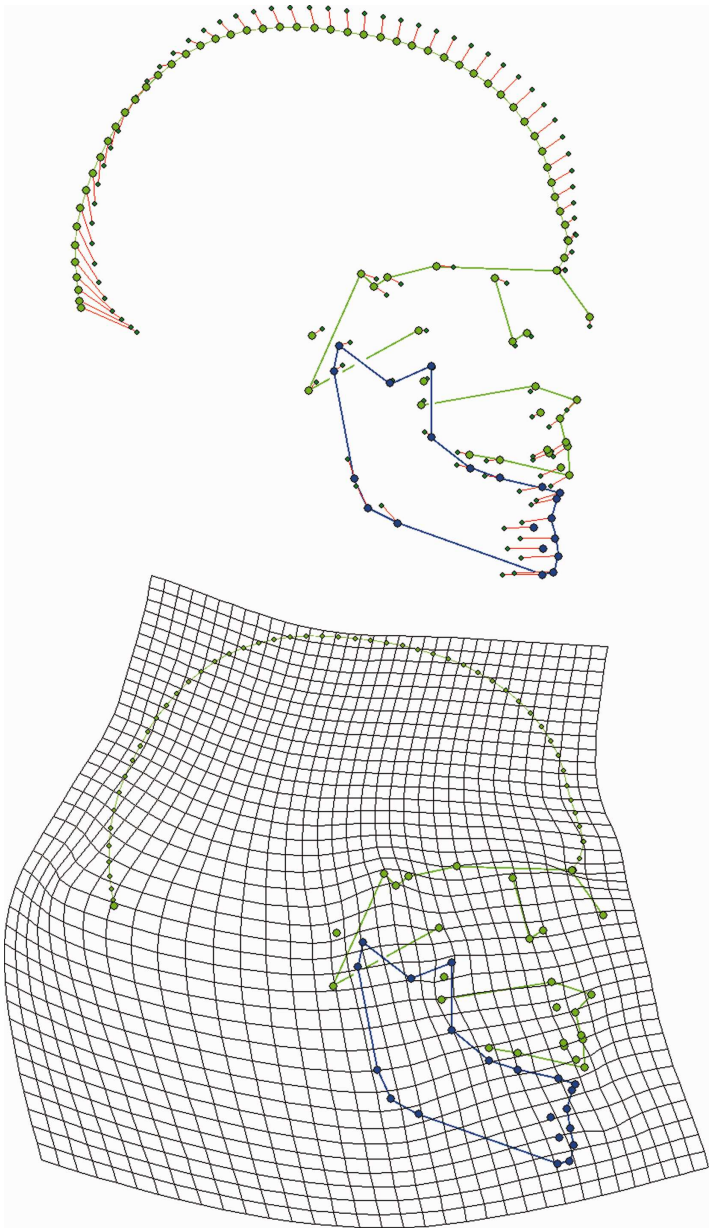

Figure 4. Two methods for depicting the same shape differences between two landmark configurations. Data were derived from lateral radiographs of modern human skulls. The images represent differences along a principal component of the sample shape variation. For clarity, mandibular landmarks are shown in blue. The upper figure illustrates differences using vectors to connect the corresponding landmarks. The lower figure uses a TPS interpolation. Options to include both configurations or a single configuration plus vectors, and choices about which landmarks to connect can greatly affect the ease with which figures are interpreted. (Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.)

wise, a distinct pattern of variation may constitute important evidence of taxonomic diversity. This approach was applied to the *H. erectus sensu lato* samples from Africa, Asia, and Eurasia, using various extant papionin and hominoid species, as well as the extinct *Theropithecus dartii*.[25] Although some of these species were more sexually dimorphic than *H. erectus* was likely to be, only the neurocranium was analyzed, which

somewhat mitigated this effect. The results were mixed: the fossil sample was more variable than some single species, but was less variable than many other species. Interestingly, the neurocranial shape of OH 9, an African specimen often assigned to the otherwise Asian *H. erectus sensu stricto*, more closely resembled other African specimens assigned to *H. ergaster*. Given the more extensive temporal and geographic range of the *H. erectus* sample, as well as the lack of a clear distinction in shape between *H. erectus s.s.* and *H. ergaster*, the null hypothesis of a single species was not rejected in this case.

Whereas the previous example used the magnitude of intraspecific variation as a yardstick to test taxonomic hypotheses, other studies have used intertaxon variation for similar purposes. To test the null hypothesis that Neanderthals were a subspecies of *H. sapiens* rather than a separate species, Harvati, Frost, and McNulty[26] compared the Mahalanobis' D^2 distances in the Neanderthal sample and several modern human populations to the distances among modern human populations and among species and subspecies of other catarrhine primates. They concluded that Neanderthals were best interpreted as a distinct species based on the observation that the distances between Neanderthals and modern humans exceeded nearly all of the intraspecific comparisons, as well as several of the interspecific comparisons.

Exploring evolutionary change in biological structure

Beyond establishing the species affinities of individual fossils and clarifying the taxonomic status of various fossil samples, workers have also explored the evolutionary implications of interspecific shape variation in the dentition,[27–29] exocranium,[30] and endocranium.[31] For example, Lieberman, McBratney, and Krovitz[30] used a combination of linear- and landmark-based analyses of the exocranium to examine the developmental processes that produce a uniquely globular neurocranium and small, retracted face in *H. sapiens* relative to archaic *Homo*. Bruner, Manzi, and Arsuaga[31] also identified two distinct allometric trajectories in endocranial shape, one that includes all extinct *Homo* species and one that is unique to *H. sapiens*, demonstrating that the large brains of Neanderthals are the end-point of an archaic allometric pattern, whereas modern humans achieved a similar brain size with a different suite of accompanying shape changes.

Identifying the underlying causes of intraspecific variation in recent humans also contributes to our understanding of hominin evolution. Cranial shape variation within modern humans has been used to address hypotheses about the diaspora of modern humans out of Africa,[32] and to examine the influence of factors such as population replacement and gene flow on the cranial morphology of prehistoric and recent human populations in the Americas.[33] Studies using GM methods have also explored variation in cranial robusticity among modern human populations to clarify both its appearance during ontogeny,[34] and its relationship to craniofacial shape and mastication.[35] For example, Baab and coworkers[35] examined covariation between discrete traits that reflect cranial robusticity and 3-D landmarks capturing overall craniofacial shape. They found a weak relationship between cranial shape and robusticity, but no relationship between robusticity and size. Both this study and an earlier caliper-based study[36] found that the most robust skulls were also long, with broader upper faces and prognathic maxillae.[36] The earlier study, however, interpreted this combination of shape features as a link between cranial robusticity and an increase in skull size.[36] Using a GM approach, the capacity to separate clearly the effects of shape versus size demonstrated that long, narrow skulls with projecting faces were robust, but not necessarily large.

The Relationship Between Size and Shape

Ontogeny and heterochrony

Some of the earliest quantitative studies in comparative anatomy were aimed at understanding developmental changes in size and shape, as well as the relationship between them (ontogenetic allometry). Within GM, allometry is generally investigated by linear regression of either superimposed coordinates or individual PCs on a size variable; the process is comparable to regression of linear measurements on a size variable such as the geometric mean. Berge and Penin's[37] analysis of ontogenetic shape change in African apes used this approach, confirming the traditional view that allometric variation is a large component of the shape differences between chimps and gorillas, but also describing how species-specific allometric and nonallometric shape variation contribute to taxonomic differences in cranial shape.

A more recent development is the analysis in size-shape space. Rather than analyzing only shape and then regressing it against size, this approach includes the natural log of centroid size as an additional variable alongside the superimposed coordinates in a PCA.[38] Because size has a larger variance than do the individual coordinates, which have been scaled to a configuration size of 1, the first PC axis will primarily reflect absolute size differences in the sample (as well as allometric shape differences that are common to the entire sample).[38] Concentrating the common allometric shape variation into a single component offers a clear advantage over a standard PCA in shape space, where this variation may be distributed across several axes.

Mitteroecker and colleagues'[38] analysis of ontogenetic allometry among great ape and human crania in size-shape space showed that humans have the most distinct cranial shape at birth, but that the orangutan cranium also differs slightly from that of the African apes even at this early stage. In addition to different initial shapes, both the human and orangutan trajectories also diverge from that of the African apes, and each other, early in ontogeny, further distinguishing these two taxa. In contrast, gorillas and chimpanzees appeared to follow very sim-

ilar trajectories in this multi-taxon analysis. However, when the analysis was limited to the African apes, it was clear that more subtle differences in the trajectories of chimpanzees and gorillas were masked in the larger ordination, and that, in agreement with Berge and Penin,[37] these taxa differ by more than just allometric scaling.

Researchers have taken advantage of juvenile fossils to extend these comparisons to extinct hominin species. For example, using the Taung child as well as adult *A. africanus* fossils, Cobb and O'Higgins[39] found that the primary direction of ontogenetic shape change in the face of *A. africanus* was more similar to that in chimpanzees and gorillas than that in modern humans (based on a comparison of the first PC in species-specific PCAs). McNulty et al.[24] obtained a similar result using the full dimensionality of shape space. They further demonstrated that despite differences between late-stage developmental patterns among African apes, any model of development applied to the Taung child predicts an adult form more like that of Sts 71 than Sts 5. Differences in ontogenetic development have been quantified even among the more closely related Neanderthals and modern humans. Based on a PCA in size-shape space, Bastir, O'Higgins, and Rosas[40] demonstrated that the mandibles of Neanderthals <1 year of age are indistinguishable from those of young modern humans, but that the species' ontogenies diverge after this time. This postnatal divergence in the ontogenetic trajectories results in the distinct adult morphologies of these two species.

In addition to the relationship between size and shape, studies of heterochrony are concerned with the association or dissociation of these factors with age. To identify the type of heterochronic process (*sensu* Gould[41]) responsible for dissociations during ontogeny, it is necessary to also analyze information about developmental stage and possibly chronological age, as well as phylogenetic information.[42,43] Heterochronic analyses using GM have typically followed traditional approaches, using bivariate plots comparing PC scores, log centroid size, and a time dimension to detect dissociations among shape, size, and age.[44,45]

An example of this approach in human paleontology is the diagnosis of rate hypermorphosis in both growth and development of the skull in Neanderthals compared to modern humans using GM methods.[44] It seems likely that species-specific morphologies were present at birth, but that differences in skull morphology were further enhanced by divergent ontogenetic trajectories and a faster postnatal rate of size and shape change in Neanderthals.[45] The Ponce de León and Zollikofer[44] study nicely demonstrates the full visual impact that is possible using 3-D GM methods by illustrating shape changes over ontogeny in modern humans and Neanderthals using color-coded vector fields superimposed on the average shape for both species (Fig. 3, p. 535, in that publication).

However, the use of bivariate comparisons of PC scores, size, and time, is somewhat controversial. The debate centers on whether similar group distributions on a limited number of PC axes (for example, PCs 1 and 2) signify that those groups undergo similar ontogenetic shape change, or whether a common allometric or ontogenetic trajectory is valid only if it can be demonstrated for the entirety of shape space (that is, along all PC axes). Researchers who advocate the use of only a limited number of PCs argue that PCs are as valid as linear measurements or ratios for describing biological shape; covariation in traits captured by PCs are likely to be the result of developmental processes; and PCs can highlight testable hypotheses regarding morphogenesis.[45] Critics of this approach point out that PCs are simply statistical constructs and do not necessarily have easy biological interpretations.[12,46,47] They also argue that similar slopes on only a few PCs does not mean that trajectories are necessarily parallel in full shape space (that is, if all PCs are inspected).[38,39,46] Moreover, individual PC axes may include shape variation not directly related to ontogeny, including interspecific shape differences.[38]

If these criticisms are correct and the signal from the first few PCs is not a reliable way to compare interspecific allometric or ontogenetic patterns, then what is the alternative? Among the possibilities, one stands out because it avoids using ordination methods like PCA altogether and instead uses direct comparison of ontogenetic or allometric trajectories within the entire shape space.[38,39] These trajectories are a list of beta (slope) coefficients for each coordinate from a multivariate regression of the coordinates on centroid size[35] or dental stages.[24] A disadvantage of this approach is that these trajectories are multivariate and cannot be visualized on a 2-D or 3-D plot. Nevertheless, one can measure the angle between these multivariate trajectories just as one measures the angle between two regression slopes in a bivariate analysis. Comparisons between groups can then proceed with reference to the angular differences.[24,35] Taken to its logical extreme, the consequence of this approach is that if the ontogenetic trajectories of two species diverge on even one PC axis, no matter how little variation that PC accounts for, we would have to reject the hypothesis of global heterochrony.

In fact, Mitteroecker, Gunz, and Bookstein[46] argue just this, based on the observation that for classic heterochronic descriptors to be valid, the underlying ontogenetic trajectories between species must be identical, thus allowing only disassociations between shape, size, and time rather than changes in shape trajectories. They demonstrated that the trajectories of *Pan troglodytes* and *P. paniscus* are not identical in full shape space, and thus rejected a purely heterochronic explanation for observed differences in craniofacial shape. Even when the cranium was divided into the neurocranium, upper face, and lower face, heterochrony for each of these regions was rejected.[46]

These results appear to contradict the findings of Lieberman and coworkers[45] who also investigated the role of heterochrony in *Pan* at the level of the entire cranium. Lieberman and coworkers' results were consistent with earlier studies based

on linear dimensions that identified heterochrony in *Pan*. They specifically identified postformation, the early underdevelopment of shape, as the likely explanation for particular shape differences (those on PC 1) between the two chimpanzee species. They also concluded that while postformation explains some important differences, particularly in the vault and base, other evolutionary processes must be invoked to explain differences in facial morphology. However, the apparent conflict between the two results is due to the fact that Lieberman and associates[45] applied a less strict criterion for heterochrony than did Mitteroecker, Gunz, and Bookstein[46] and accepted an apparent pattern of heterochrony on PC 1 as biologically meaningful even if the trajectories did not overlap on all PCs.

Ultimately, resolution of this disagreement is not simple. On one hand, the argument that ontogenetic trajectories should be identical before invoking heterochronic explanations is reasonable. On the other hand, it may be relatively easy to find a significant difference between two multi-dimensional vectors, given the copious amounts of data that can now be collected and analyzed in GM analyses. Unfortunately, there is no clear rule for how many variables must differ and by how much before two trajectories are "significantly" different (in a biological rather than a statistical sense). While heterochrony provides an elegant theoretical framework for explaining evolutionary change in morphology, it seems increasingly unlikely that global heterochrony will be validated in multivariate analyses of shape. Rather than focusing solely on significant *p*-values, therefore, it may be more fruitful to identify the actual similarities and differences among shape trajectories, assess the timing of those differences, and track their ultimate impact on adult morphology. These avenues of investigation have the potential to identify new hypotheses related to the developmental and evolutionary pathways involved in lineage diversification.

Modularity and Integration

The idea that biological processes such as heterochrony act differentially across a structure is intimately tied to the concept of modularity. Modularity predicts that developmental, functional, or other biological processes produce natural partitions among sets of characters, called modules. For example, pioneering work by Moss[48] emphasized the importance of epigenetic effects from the surrounding soft tissues and spaces (functional matrices) during development of the distinct cranial components. A key assumption of current modularity studies is that traits exhibit stronger covariation within than among modules.[49] The weaker among-module connectivity may allow for the semi-independent action of evolutionary and developmental processes (such as differential heterochrony) on individual modules. Conceptually, modularity is related to morphological integration, which is based on the idea that individual morphological features do not evolve in isolation; rather, individual phenotypic features evolve in a coordinated fashion.

Analytical methods related to integration and modularity have several goals, including quantification of the magnitude of integration (for example, among modules or taxa), identification of the underlying process leading to integration, verification of *a priori* defined modules, and comparing and visualizing patterns of integration or modularity.[50,51] We will focus on the latter two points.

Verification of *a priori* modules

Several approaches are available to test whether empirical data are congruent with hypothesized modules. One approach involves comparing the correlation matrix from the empirical measurements to a connectivity matrix of expected correlations based on the hypothesized modules (derived from genetic, developmental, or functional considerations).[51,52] In this context, the connectivity matrix is composed of a "1" or a "0" in the cell corresponding to each pair of traits, depending on whether or not the traits are expected to be in the same module. The matrix correlation between the original, empirically derived correlation matrix and the connectivity matrix is used to assess whether the observed pattern of trait correlations corresponds to the hypothesized modules. This approach has been applied broadly across mammals[53] and within primates,[54–57] including hominoids,[58] using linear measurements. An alternative approach is to calculate the correlation (or covariation) between the traits assigned to each hypothesized module.[51,59,60] This value can then be compared to an empirical distribution of correlation values generated through a permutation procedure that randomly assigns traits, without replacement, to modules. If the hypothesized modules are indeed true biological modules, then the correlation between the hypothesized modules should be relatively low, reflecting their (semi-) independence from one another.[51,59,60]

A potential challenge to both of these methods is the fact that a spatially structured pattern of covariance among landmarks could masquerade as evidence for modules. For example, the strength of covariation between any two landmarks could be negatively correlated with their interlandmark distances. This is a reasonable expectation because in many cases "internal integration of modules relies on tissue-bound interactions among their parts." [51:412] However, due to the highly uneven distribution of landmarks on most biological structures, spatially disjunct clusters of landmarks would behave as distinct modules regardless of any underlying biological modularity. In most studies of modularity, each *a priori* module is defined as a set of traits from a restricted morphological region (for example, the bones surrounding the nasal cavity) rather than landmarks from spatially disparate regions (for example, the bones surrounding the nasal cavity and the foramen magnum). This presents a distinct challenge because it means that low levels of integration among *a-priori*-defined modules is not necessarily support for the underlying theoretical model on which these partitions were based. This pattern of spatial autocorrelation may explain why many studies

find support for multiple alternative models of modularity.[61,62]

One way to address this problem is to modify the second test described above to require that the landmarks assigned to each module during the permutation process be spatially contiguous (that is, neighboring landmarks).[51,58,59] This more closely resembles the reality that hypothesized modules are composed of landmarks from a restricted spatial region and therefore provide a more conservative test of modularity. However, it remains unclear whether even this solution goes far enough, as suggested by the two examples presented by Klingenberg[51] to demonstrate this method. This permutation test rejected the anterior and posterior compartments of the fly wing as distinct modules, but supported the presence of ascending ramus and alveolar modules in the mouse mandible.[51] However, the proposed modules of the wing are not nearly as spatially disjunct as are the modules proposed for the mouse mandible. Furthermore, there is no other combination of the mouse mandible landmarks that results in two modules as spatially isolated as the hypothesized modules. It is therefore difficult to know whether these results should be interpreted as supporting the developmental modularity of the mouse mandible but not the fly wing, or whether this simply reflects a pattern of spatial autocorrelation among landmarks in both structures.

The increasing popularity of modularity studies in anthropology makes it likely that these approaches will soon be extended to hominins in a geometric morphometric framework. We suggest that it may be advisable to establish that a simple spatial model of covariation is insufficient to explain the observed covariation patterns before testing whether theoretically derived *a priori* modules are supported by empirical data. The means and criteria for doing so, however, have not yet been established. Of course, even if the observed pattern of covariation is consistent with a simple spatial model, this does not rule out the presence of developmental or functional modularity. The difficulty in distinguishing between these two interpretations, however, does highlight the need for more sensitive tests of hypothesized modules.

Comparing and visualizing patterns of integration or modularity

The GM toolkit offers a distinct advantage over linear morphometrics in its ability to visualize coordinated shape change (that is, integration) across a structure or among structures. For example, two-block partial least squares analysis (2-B PLS), also called singular warps analysis, is frequently used to visualize covariation between different structures or anatomical regions. 2-B PLS calculates the direction of maximum covariation shared between two blocks of variables, such as the face and the neurocranium, and each successive pair of axes accounts for progressively less of the shared covariance (much as successive PC axes account for progressively less of the total variance).[17,52,53] Each specimen can be plotted along the corresponding axes, one axis for each block of variables, and the coordinated shape variation captured by these axes can be visualized.

A 2-B PLS analysis of 3-D cranial landmarks subdivided into facial, neurocranial, and basicranial regions found that the main patterns of integration among these cranial regions are shared across taxa and throughout ontogeny in African apes and humans.[54,55] In fact, a large proportion of the shape differences among these taxa is due to differential truncation or extension of shared developmental pathways, with a smaller proportion of the intertaxic shape differences due to more modular (local) changes. These results confirmed an earlier analysis of integration in African apes and humans based on linear measurements.[54] A recent investigation further confirmed that *Pongo* also shares the primary patterns of integration with African apes and humans, despite its distinct cranial shape.[56] These comparisons are important for understanding the evolution of developmental pathways and may explain why evolution appears to occur more frequently in certain directions than others. Furthermore, this work has bearing on the use of living taxa as analogs for extinct species; although similar, patterns of integration in living hominoids are not identical to one another, and are unlikely to be identical in extinct species.[54]

In regard to integration in fossil hominins, Bookstein and colleagues[17] extended the two-block PLS to a three-block model, using one of several possible algorithms, to examine integration among the cranial vault, base, and face simultaneously rather than on a pair-wise basis. Careful sample design allowed covariation among cranial regions to be investigated both over ontogeny in *H. sapiens* and through evolution in Pleistocene *Homo*. Overall, the patterns of developmental and evolutionary integration were similar, with the largest divergence occurring in the cranial base: the anterior cranial base underwent a rotation during evolution that is not mirrored by ontogeny. Using the same three-block approach, Gunz and Harvati[57] found that modern humans and Neanderthals have similar patterns of integration among the temporal bones, midsagittal parietal, and midsagittal occipital profiles. In both taxa, an occipital bun occurred along with a flattened midsaggital parietal profile and with temporal bones positioned more anterosuperiorly. In this context, the Neanderthal pattern of extreme "bunning" could be viewed as an extrapolation of the human pattern of cranial integration. What remains to be clarified, in this and other similar studies, is what underlying forces result in these patterns of covariation.

FUTURE DIRECTIONS AND CHALLENGES

Having highlighted the most common and, arguably, successful applications of GM approaches to paleoanthropological questions, we turn to several areas that are only beginning to be explored and represent both

great promise and challenge for novel applications of GM analyses: fossil reconstruction, functional morphology, quantitative genetic models, and phylogeny reconstruction. Some workers have already begun exploring these areas, but additional work remains in all cases.

Fossil Reconstruction

The combination of visual and statistical tools represented by the GM toolkit has enormous potential for aiding in the reconstruction of damaged and warped fossil specimens. While fossil reconstruction using these methods is already well established in some research teams, the complex algorithms and expensive software used for extensive reconstructive "surgery" have made this application the domain of very few practitioners. More widely used are the less technically sophisticated methods for estimating the position of missing landmarks based on the existing ones. The most straightforward case is in a symmetrical specimen when a bilateral landmark is missing on only one side. The landmark can be "mirrored" from the preserved to the missing side in a process called reflected relabeling, which exploits the inherent symmetry of biological organisms.[23,58] Other options exist for estimating midline landmarks or bilateral landmarks missing on both sides, including using the average landmark position in a morphologically similar comparative sample or interpolating the position of a missing landmark based on multiple regression or a thin-plate spine function.[58]

More elaborate and detailed reconstructions of fossil crania include repositioning displaced or broken morphology, unwarping areas that have suffered plastic deformation, and estimating the morphology of missing surfaces. These reconstructions allow for more robust comparisons across a fragmented fossil record and also facilitate research using methods such as finite element analysis. Several studies that incorporate complex reconstructions have now been published,[18,58,59] and Zollikofer and Ponce de León[60] have published a book describing methods for virtual reconstructions. We refer interested readers to these sources for details of these methods. Nevertheless, the future potential of this application of GM is in making virtual reconstruction more accessible to the paleoanthropology community in terms of data acquisition (availability of digital fossil representations) and software (priced toward modest anthropology budgets).

Functional Morphology

More functionally oriented applications of GM have either compared shape differences to *a priori* predictions based on theoretical functional models[72] or done a 2-B PLS analysis of shape variables and functional variables.[73] Although the linear and angular measurements that are the mainstay of comparative functional morphology fail to account for the complex 3-D geometry of biological structures, they have been chosen to maximize the (functional) signal-to-noise ratio. In contrast, landmarks provide a more realistic representation of 3-D shape, but it is difficult to separate the functional signal from the other signals also encoded in the landmark data, such as, phylogenetic history. Choosing a landmark configuration that reflects the most functionally relevant aspects of shape, as well as statistical approaches that can partition out the different sources of variation, may be useful in overcoming this obstacle. Furthermore, coordinate data may be particularly useful in detecting morphological changes that are relatively subtle, such as those between groups subject to different experimental conditions. As an example, animals raised on different diets may exhibit slight but consistent osteological changes more easily captured by 3-D landmark data than by traditional measures.

An innovative application of GM was the evaluation of the effect of including the periodontal ligament in a finite element model of the human mandible by measuring shape change associated with deformation alongside the traditional analysis of strain data.[61,62] This study represented an application of methods for combining geometric morphometrics and functional simulations first developed by O'Higgins and coworkers.[61] In addition to changes associated with including the periodontal ligament (versus no ligament), they also documented different patterns of shape change in models where varying material properties were assigned to the ligament. It is also possible to use coordinate data to track changes in posture by collecting landmarks (at joint positions, for example) on the same individual at consecutive time slices. This ordered sequence of landmark configurations (a trajectory) contains information about changes in position over the course of a movement. By partitioning this trajectory into shape, size, and orientation components, it becomes possible to quantitatively describe and compare different motions or different individuals performing the same motion.[63]

Quantitative Genetics

One emerging field of interest is the study of how quantitative genetic models can explain morphological variation. Quantitative genetics provides insights into the heredity of continuous traits and the underlying evolutionary processes, such as selection or genetic drift, that generate and maintain variation.[64] In quantitative genetics, parameters such as heritability are estimated from large pedigreed samples or breeding experiments. The existence of modern human cranial data sources with known or estimated pedigrees has allowed anthropologists to study the heredity of craniofacial morphology. One of these sources, the FELS Longitudinal Study, includes data from more than a thousand modern humans, including extended families with known relationships and DNA sequences for many individuals. A study conducted by Sherwood and McNulty[65] used landmarks and semilandmarks to estimate heritability in craniofacial shape and correlations between components of shape variation and genetic variance. The results were highly suggestive, demonstrating that specific regions of the ge-

nome may be linked to particular variations in cranial shape, including variations in cranial flexion, midfacial prognathism, chin development, and other features that figure prominently in hypotheses about human evolution. A larger study to test these linkages and to look for additional correlations is currently under way. Further research along these lines will be possible as research on the human, chimpanzee, and Neanderthal genomes leads to additional insights into the genotype-to-phenotype map for hominins.

Phylogeny Estimation

Reconstructing phylogeny with shape data is of great interest within the paleoanthropology community, but has proven challenging. There have been some limited attempts to use landmark-derived shape data to estimate phylogenetic relationships based on converting shape information into discrete characters,[66,67] using phylogenetic methods developed specifically for continuous data,[68] or tree-building algorithms that cluster based on distance measures.[69] The long-standing debate over the theoretical and empirical issues surrounding the conversion of continuous traits to discrete characters notwithstanding, the recent method advocated by González-José et al.[67] has been strongly criticized on methodological grounds[70] and should not be applied.

Lockwood, Kimbel, and Lynch[69] recovered the "correct" phylogenetic relationships (that is, those based on genetic data) among modern humans and apes by applying neighbor-joining and ordinary least squares algorithms to Procrustes distances from 3-D temporal bone landmarks. They attributed their positive results to both the repeatability and higher resolution of landmark data compared to qualitative and traditional morphometric data, and to the fact that the temporal bone's "functional complexity should minimize the possibility that a single behavioral shift in unrelated taxa could lead to homoplastic similarity across a suite of features."[69:4356] However, this study was limited to the large-bodied apes, and it is unclear whether these results could be replicated with a larger primate sample. Furthermore, subsequent studies have found that the parietal, sphenoid, and frontal bones are just as strongly correlated as is the temporal bone with neutral genetic distances among modern human populations.[71] Overall cranial vault shape has also been shown to correlate with neutral genetic distances among human populations.[72]

Together, these results indicate that information regarding genetic relationships may be encoded in cranial shape, but there is not yet a reliable way to separate the phylogenetic signal from the homoplastic information. For example, patterns of allometry can manifest in distinct lineages as convergent morphology. Simple methods to account for this, such as by regressing shape variables against log centroid size, may be inappropriate in many circumstances. Other forces besides size changes can lead to homoplasies (for example, functional convergence), and it is also unclear how to account for these. These difficulties are not unique to GM or even traditional quantitative variables used in phylogeny reconstruction. However, when adopting a quantitative approach it becomes necessary to account for such factors in a rigorous and repeatable fashion. A final obstacle to using landmark data in phylogeny reconstruction is that there is no single "best" solution for extracting the phylogenetic information. Rohlf[68] suggested the use of phylogenetic methods designed for continuous data, particularly the maximum likelihood methods of Felsenstein,[73] but this has not been used in practice.

CONCLUSIONS

The GM toolkit evolved through a blending of shape theory from the field of mathematics, thin-plate spline analogy from engineering, and multivariate analysis championed by biostatisticians. This synthesis resulted in a biologically and statistically powerful tool for studying form that has substantially affected the study of human evolution. We have highlighted some of the contributions to date. However, the full power of GM will not be realized until further integration takes place between GM and analytical approaches from functional morphology, quantitative genetics, phylogenetic comparative methods, and phylogeny estimation. Important strides forward are being made in these areas, both from within biological anthropology and from closely allied disciplines like evolutionary biology. We see great potential for GM to contribute further to paleoanthropological research, not only in the areas discussed here, but also in new applications and modes of analysis.

ACKNOWLEDGMENTS

We thank John Fleagle for his initial suggestion that we write this review, and for his continued patience. We also thank the reviewers for their insightful and helpful comments on this manuscript.

REFERENCES

1 Howells WW. 1973. Cranial variation in man. Papers of the Peabody Museum of Archaeology and Ethnology 67. Cambridge: Harvard University Press.

2 Oxnard CE. 1973. Form and pattern in human evolution. Chicago: University of Chicago Press.

3 Rohlf FJ, Marcus LF. 1993. A revolution in morphometrics. Trends Ecol Evol 8:129–132.

4 Bookstein FL. 1991. Morphometric tools for landmark data: geometry and biology. Cambridge: Cambridge University Press.

5 Adams DC, Rohlf FJ, Slice DE. 2004. Geometric morphometrics: ten years of progress following the "revolution." Ital J Zool 71:5–16.

6 Slice DE. 2007. Geometric morphometrics. Annu Rev Anthropol 36:261–281.

7 Bookstein FL. 1997. Landmark methods for forms without landmarks: localizing group differences in outline shape. Med Image Anal 1:225–243.

8 Perez SI, Valeria B, Paula NG. 2006. Differences between sliding semi-landmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. J Anat 208:769–784.

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

10 Rohlf FJ, Slice D. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst Zool 39:40–59.

11 Kent JT. 1994. The complex Bingham distribution and shape analysis. J Roy Stat Soc B 56:285–299.

12 Rohlf FJ. 1999. Shape statistics: Procrustes superimpositions and tangent spaces. J Classif 16:197–223.

13 Kendall DG. 1989. A survey of the statistical theory of shape. Stat Sci 4:87–99.

14 Slice DE. 2001. Landmark coordinates aligned by Procrustes analysis do not lie in Kendall's shape space. Syst Zool 50:141–149.

15 Marcus LF, Hingst-Zaher E, Zaher H. 2000. Application of landmark morphometrics to skulls representing the orders of living mammals. Hystrix 11:27–47.

16 Bookstein FL. 1996. Biometrics, biomathematics and the morphometric synthesis. Bull Math Biol 58:313–365.

17 Bookstein FL, Gunz P, Mitteroecker P, et al. 2003. Cranial integration in *Homo*: singular warps analysis of the midsagittal plane in ontogeny and evolution. J Hum Evol 44:167–187.

18 Zollikofer CPE, Ponce de Leon MS, Lieberman DE, et al. 2005. Virtual cranial reconstruction of *Sahelanthropus tchadensis*. Nature 434:755–759.

19 Guy F, Lieberman DE, Pilbeam D, et al. 2005. Morphological affinities of the *Sahelanthropus tchadensis* (Late Miocene hominid from Chad) cranium. Proc Natl Acad Sci USA 102:18836–18841.

20 Grine FE, Bailey RM, Harvati K, et al. 2007. Late Pleistocene human skull from Hofmeyr, South Africa, and modern human origins. Science 315:226–229.

21 Delson E, Harvati K, Reddy D, et al. 2001. The Sambungmacan 3 *Homo erectus* calvaria: a comparative morphometric and morphological analysis. Anat Rec 262:380–397.

22 Baab KL. 2008. A re-evaluation of the taxonomic affinities of the early *Homo* cranium KNM-ER 42700. J Hum Evol 55:741–746.

23 Baab KL, McNulty KP. 2009. Size, shape, and asymmetry in fossil hominins: the status of the LB1 cranium based on 3D morphometric analyses. J Hum Evol 57:608–622.

24 McNulty KP, Frost SR, Strait DS. 2006. Examining affinities of the Taung child by developmental simulation. J Hum Evol 51:274–296.

25 Baab KL. 2007. Cranial shape variation in *Homo erectus*. PhD dissertation, City University of New York.

26 Harvati K, Frost SR, McNulty KP. 2004. Neanderthal taxonomy reconsidered: implications of 3D primate models of intra-and interspecific differences. Proc Natl Acad Sci USA 101:1147–1152.

27 Gómez-Robles A, Martinón-Torres M, Bermúdez de Castro JM, et al. 2007. A geometric morphometric analysis of hominin upper first molar shape. J Hum Evol 53:272–285.

28 Gómez-Robles A, Martinón-Torres M, Bermúdez de Castro JM, et al. 2008. Geometric morphometric analysis of the crown morphology of the lower first premolar of hominins, with special attention to Pleistocene *Homo*. J Hum Evol 55:627–638.

29 Martinón-Torres M, Bastir M, Bermúdez de Castro JM, et al. 2006. Hominin lower second premolar morphology: evolutionary inferences through geometric morphometric analysis. J Hum Evol 50:523–533.

30 Lieberman DE, McBratney BM, Krovitz G. 2002. The evolution and development of cranial form in *Homo sapiens*. Proc Natl Acad Sci USA 99:1134–1139.

31 Bruner E, Manzi G, Arsuaga JL. 2003. Encephalization and allometric trajectories in the genus *Homo*: evidence from the Neandertal and modern lineages. Proc Natl Acad Sci USA 100:15335–15340.

32 Gunz P, Bookstein FL, Mitteroecker P, et al. 2009. Early modern human diversity suggests subdivided population structure and a complex out-of-Africa scenario. Proc Natl Acad Sci USA 106:6094–6098.

33 Perez SI, Bernal V, Gonzalez PN. 2009. Evolutionary relationships among prehistoric human populations: an evaluation of relatedness patterns based on facial morphometric data using molecular data. Hum Biol 79:25–50.

34 Gonzalez PN, Perez SI, Bernal V. 2010. Ontogeny of robusticity of craniofacial traits in modern humans: a study of South American populations. Am J Phys Anthropol 142:367–379.

35 Baab KL, Freidline SE, Wang SL, et al 2010. Relationship of cranial robusticity to cranial form, geography and climate in *Homo sapiens*. Am J Phys Anthropol 141:97–115.

36 Lahr MM, Wright RVS. 1996. The question of robusticity and the relationship between cranial size and shape in *Homo sapiens*. J Hum Evol 31:157–191.

37 Berge C, Penin X. 2004. Ontogenetic allometry, heterochrony, and interspecific differences in the skull of African apes, using tridimensional Procrustes analysis. Am J Phys Anthropol 124:124–138.

38 Mitteroecker P, Gunz P, Bernhard M, et al. 2004. Comparison of cranial ontogenetic trajectories among great apes and humans. J Hum Evol 46:679–698.

39 Cobb SN, O'Higgins P. 2004. Hominins do not share a common postnatal facial ontogenetic shape trajectory. J Exp Zool B 302:302–321.

40 Bastir M, O'Higgins P, Rosas A. 2007. Facial ontogeny in Neanderthals and modern humans. Proc R Soc B 274:1125–1132.

41 Gould SJ. 1977. Ontogeny and phylogeny. Cambridge: Harvard University Press.

42 Godfrey L, Sutherland MR. 1996. Paradox of peramorphic paedomorphosis: heterochrony and human evolution. Am J Phys Anthropol 99:17–42.

43 Klingenberg CP. 1998. Heterochrony and allometry: the analysis of evolutionary change in ontogeny. Biol Rev Camb Philos Soc 73:79–123.

44 Ponce de León MS, Zollikofer CPE. 2001. Neanderthal cranial ontogeny and its implications for late hominid diversity. Nature 412:534–538.

45 Lieberman DE, Carlo J, Ponce de León M, et al. 2007. A geometric morphometric analysis of heterochrony in the cranium of chimpanzees and bonobos. J Hum Evol 52:647–662.

46 Mitteroecker P, Gunz P, Bookstein FL. 2005. Heterochrony and geometric morphometrics: a comparison of cranial growth in *Pan paniscus* versus *Pan troglodytes*. Evol Dev 7:244–258.

47 Mitteroecker P, Gunz P, Weber GW, et al. 2004. Regional dissociated heterochrony in multivariate analysis. Ann Anat 186:463–470.

48 Moss ML. 1972. Twenty years of functional cranial analysis. Am J Orthodont 61:479–485.

49 Winther RG. 2001. Varieties of modules: kinds, levels, origins, and behaviors. J Exp Zool 291:116–129.

50 Chernoff B, Magwene PM. 1999. Afterword: morphological integration: forty years later. In: Olson EC, Miller RL, editors. Morphological integration. Chicago: University of Chicago Press. p 319–353.

51 Mitteroecker P, Bookstein F. 2007. The conceptual and statistical relationship between modularity and morphological integration. Syst Biol 56:818–836.

52 Bastir M, Rosas A. 2005. Hierarchical nature of morphological integration and modularity in the human posterior face. Am J Phys Anthropol 128:26–34.

53 Rohlf FJ, Corti M. 2000. Use of two-block partial least-aquares to study covariation in shape. Syst Biol 49:740–753.

54 Ackermann RR. 2002. Patterns of covariation in the hominoid craniofacial skeleton: implications for paleoanthropological models. J Hum Evol 43:167–187.

55 Mitteroecker P, Bookstein FL. 2008. The evolutionary role of modularity and integration in the hominoid cranium. Evolution 62:943–958.

56 Singh N, Harvati K, Hublin J-J, et al. 2012. Morphological evolution through integration: a quantitative study of cranial integration in *Homo*, *Pan*, *Gorilla* and *Pongo*. J Hum Evol 62:155–164.

57 Gunz P, Harvati K. 2007. The Neanderthal "chignon": variation, integration, and homology. J Hum Evol 52:262–274.

58 Gunz P, Mitteroecker P, Neubauer S, et al. 2009. Principles for the virtual reconstruction of hominin crania. J Hum Evol 57:48–62.

59 Grine FE, Gunz P, Betti-Nash L, et al. 2010. Reconstruction of the Late Pleistocene human skull from Hofmeyr, South Africa. J Hum Evol 59:1–15.

60 Zollikofer CPE, Ponce de León MS. 2005. Virtual reconstruction: a primer in computer-assisted paleontology and biomedicine. New York: Wiley.

61 O'Higgins P, Cobb SN, Fitton LC, et al. 2011. Combining geometric morphometrics and functional simulation: an emerging toolkit for virtual functional analyses. J Anat 218:3–15.

62 Gröning F, Fagan MJ, O'Higgins P. 2011. The effects of the periodontal ligament on mandibular stiffness: a study combining finite element analysis and geometric morphometrics. J Biomech 44:1304–1312.

63 Adams DC, Cerney MM. 2007. Quantifying biomechanical motion using Procrustes motion analysis. J Biomech 40:437–444.

64 Lande R. 1979. Quantitative genetic-analysis of multivariate evolution, applied to brain - body size allometry. Evolution 33:402–416.

65 Sherwood RJ, McNulty KP. 2011. Dissecting the genetic architecture of craniofacial shape. In: Lestrel PE, editor. Biological shape analysis. Singapore: World Scientific. p 145–171.

66 Bookstein FL. 2000. Creases as local features of deformation grids. Med Image Anal 4:93–110.

67 González-José R, Escapa I, Neves WA, et al. 2008. Cladistic analysis of continuous modularized traits provides phylogenetic signals in *Homo* evolution. Nature 453:775–778.

68 Rohlf FJ. 1998. On applications of geometric morphometrics to studies of ontogeny and phylogeny. Syst Biol 47:147–158.

69 Lockwood CA, Kimbel WH, Lynch JM. 2004. Morphometrics and hominoid phylogeny: support for a chimpanzee-human clade and differentiation among great ape subspecies. Proc Natl Acad Sci USA 101:4356–4360.

70 Adams DC, Cardini A, Monteiro LR, et al. 2011. Morphometrics and phylogenetics: principal components of shape from cranial modules are neither appropriate nor effective cladistic characters. J Hum Evol 60:240–243

71 von Cramon-Taubadel N. 2009. Congruence of individual cranial bone morphology and neutral molecular affinity patterns in modern humans. Am J Phys Anthropol 140:205–215.

72 Harvati K, Weaver TD. 2006. Human cranial anatomy and the differential preservation of population history and climate signatures. Anat Rec 288A:1225–1233.

73 Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376.

74 Dryden IL, Mardia KV. 1998. Statistical shape analysis. New York: John Wiley and Sons.