SUPPLEMENTARY ONLINE MATERIAL

for

A PRACTICAL INTRODUCTION TO LANDMARK-BASED GEOMETRIC MORPHOMETRICS

MARK WEBSTER

Department of the Geophysical Sciences, University of Chicago, 5734 South Ellis Avenue, Chicago, IL 60637 <mwebster@geosci.uchicago.edu>

and

H. DAVID SHEETS

Department of Physics, Canisius College, 2001 Main Street, Buffalo, NY 14208

<sheets@canisius.edu>

THE MAIN paper introduces the exploratory and confirmatory techniques commonly used in geometric morphometrics. A practical guide showing how these methods can be performed in free morphometrics software is presented in the appendix to the main paper.

This supplement contains the mathematical details of those methods. Additional, less commonly used, superimposition methods and the theory of shape are also discussed here.

SEMILANDMARKS

Semilandmarks were briefly introduced in the main paper. The procedure for calculating semilandmark coordinates is summarized below. There are two general approaches to data acquisition: curve tracing, and the use of a template ("fan"). The two approaches appear to produce similar results (Sheets et al., 2004, 2006) and we present here only the simpler curve-tracing method. The initial steps can be performed in the tpsDig software (Rohlf, 2009). First, click the "draw background curves" tool and digitize points along the curve of interest. Use the zoom feature to ensure that all points lie on the curve: the position of any point can be adjusted by dragging it. The number of points digitized at this step is arbitrary, but the more the better. The spacing of points along the curve will likely be uneven, determined by the complexity of the curve. Next, decide how many evenly spaced points will be used to summarize the form of the curve. There is no simple rule for doing this. The denser the sampling of the curve, the more accurate is the summary of the curve following the sliding procedure (below). However, too many points can introduce excessive redundancy of information. Different curves on the same specimen need not have the same number of points, but homologous curves on all specimens must. The number of points along any curve should therefore be determined by the maximal complexity of that curve within the sample. Right click on curve and use the "resample curve" option to obtain the chosen number of evenly spaced points. Repeat

this procedure for all curves on all specimens. Save the data (this will generate a *TPS* format file that includes curves as well as landmarks).

The next step is to slide the points until their spacing minimally impacts the amount of shape difference between forms. The coordinates of these optimally slid points (semilandmarks) will be incorporated into the final configuration. Semilandmarks slide along an axis parallel to the line between the two adjacent semilandmarks (hence the desire to densely sample the curve). Two optimization criteria have been proposed (Green, 1996; Sampson et al., 1996; Bookstein, 1997; Andresen et al., 2000; Bookstein et al., 2002; Gunz et al., 2005): (1) to minimize the bending energy between each form and the mean form (Text Box 3; above); or (2) to minimize the Procrustes distance between each form and the mean form (Text Box 2; above). Both of these criteria can be supported on theoretical grounds, and choice of which to use is unlikely to affect the final inference of an analysis, although the bending energy approach does increase the variance present in the data set. For either criterion, the sliding of semilandmarks is an iterative procedure: (1) all configurations are placed in Procrustes superimposition; (2) the mean (consensus) form is calculated; (3) the bending energy or Procrustes distance of each specimen from the mean form is calculated; (4) each semilandmark on each specimen is slid along the curve to reduce the bending energy or Procrustes distance away from the mean form (the location of standard landmarks is not adjusted); and (5) steps 1-4 are iteratively repeated until the mean form no longer changes. The semilandmarks are now optimally spaced.

This sliding procedure can be conducted using the tpsRelW software (Rohlf, 2008) or the IMP SemiLand6 program. The tpsRelW program offers options for either

the bending energy criterion or the Procrustes distance criterion to optimize sliding semilandmark placement during superimposition. The former criterion has not been implemented in the SemiLand6 program, which offers only a Procrustes distance optimization criterion. SemiLand6 does allow reduction of the number of points used to represent each curve, so that the effects of varying the point density along a curve on results can be examined. Each program requires input of both a data file (landmarks plus curve data) and a protocol file. The format of the protocol files differs between the programs, but basically specifies which points in the data file are landmarks and which are semilandmarks. Details of the protocol file format and operation of each program are provided in the help file or PDF manual that accompanies each program. (Note that SemiLand6 will not accept the *TPS* data file straight from tpsDig. The data file must first be edited by deleting all "curves=" and "points=" lines and hard returns, so that just header, x- and y-coordinates, scale, and specimen identifier remain. This file must then be run through CoordGen and converted into the *X1, Y1,… CS* format, which can be loaded into SemiLand6.) Output files from tpsRelW or SemiLand6 contain coordinates of both landmarks and semilandmarks, and can be loaded into other morphometric software just like standard landmark data files.

PROCRUSTES SUPERIMPOSITION METHODS

Partial Procrustes Superimposition.—The basic principle of partial Procrustes superimposition is straightforward: the configuration of landmarks corresponding to one specimen (the target form) is matched to that of a second specimen (the reference form). In order to standardize the entire procedure the reference form is rescaled to have unit

centroid size and translated to have a centroid position of zero. The target configuration is moved (translated), rescaled, and rotated to bring it into optimal alignment with reference form. These operations do not affect the shape of the target configuration (following Kendall's [1977] definition of shape). Procrustes methods use a least squares criterion for optimizing the alignment, in which the summed squared differences in landmark positions between the target and reference is minimized. The minimization criteria can be readily derived and are presented here.

Translation: A configuration is centered by subtracting the values of the x- and y- (and z-) coordinates of the centroid from the value of the x- and y- (and z-) coordinates of each landmark in the configuration. The effect is to translate the whole configuration so that its centroid now has the coordinates [0, 0] (or [0, 0, 0]). When both reference and target forms are centered, difference in location is removed.

Rescaling: The centroid size of configuration A can be found as:

$$\text{Centroid Size (A)} = \sqrt{\sum_{i=1 \text{ to } k} \sum_{j=1 \text{ to } m} (a_{ij} - c_j)^2}$$

for k landmarks in m dimensions, where a_{ij} is the landmark j-coordinate of landmark i in configuration A, and c_j is the centroid j-coordinate. Scaling to unit centroid size is achieved by dividing the x- and y- (and z-) coordinate values of each landmark in the configuration by the centroid size of that configuration, so that:

$$\text{Rescaled Centroid Size (A)} = \sqrt{\sum_{i=1 \text{ to } k} \sum_{j=1 \text{ to } m} (a_{ij} - c_j)^2} = 1$$

When both reference and target forms are scaled to unit centroid size, difference in size is removed.

Rotation: For 2D data the angle of rotation (θ) that brings the target configuration T into optimal alignment with the reference configuration R is calculated as:

$$\theta = \text{arctangent}\left(\frac{\sum_{j=1\ \text{to}\ k} y_{Rj}x_{Tj} - x_{Rj}y_{Tj}}{\sum_{j=1\ \text{to}\ k} x_{Rj}x_{Tj} + y_{Rj}y_{Tj}}\right)$$

where x_j and y_j are the x- and y-coordinates, respectively, of landmark *j* on either the target ($_{Tj}$) or reference ($_{Rj}$) configuration. (The procedure for finding this optimal alignment for 3D data is somewhat more complicated, and is given by Rohlf [1990].)

If more than two configurations are to be aligned, an iterative procedure is followed: after centering and rescaling the configurations, (1) each configuration is sequentially rotated into optimal alignment with the first configuration using the above formula; (2) an average (consensus) configuration is calculated from these aligned configurations; (3) this consensus configuration is then used as the reference configuration in a second round of rotation of all configurations. Steps (2) and (3) are repeated until the newly calculated consensus configuration does not differ from the previously calculated consensus configuration to within some specified level of tolerance, typically far below the measurement resolution of the system. This process of iteratively estimating the mean and superimposing all specimens on it is called a Generalized Procrustes Analysis (GPA; Gower, 1975; Rohlf and Slice, 1990; Dryden and Mardia, 1998).

Full Procrustes Superimposition.—Full Procrustes superimposition is identical to partial Procrustes superimposition except that centroid sizes of the target specimen are allowed to vary, rather than being fixed at 1. This alteration allows the summed squared interlandmark distances to be adjusted to find the true minimal distance between configurations. This results in full Procrustes superimposition at the full Procrustes

distance (D_F) rather than partial Procrustes distance (D_p; appearing when centroid size is fixed at 1 [below]).

Full Procrustes superimposition places configurations in Kendall's shape space (below). Both full and partial Procrustes superimpositions result in $km - m - 1 - \{m(m - 1)/2\}$ degrees of freedom (d.f.; independent measurements) remaining in the data. For 2D data there are therefore km free variables but only $2k - 4$ d.f. following full Procrustes superimposition ($3k - 7$ for 3D data). The loss of d.f. results from the superimposition process, but is not obvious from looking at the resulting landmark positions after the superimposition because the number of variables remains unchanged. This feature of landmark data can cause software intended for use with standard morphometric data to incorrectly compute d.f. values when carrying out statistical tests.

Although full Procrustes superimposition finds the true minimal distance between configurations, the partial Procrustes method is often the preferred choice for superimposition because it produces distances between specimens (see below) that more closely match the distances in the curved, non-Euclidean spaces that most accurately describes the spacing of specimens relative to one another (Dryden and Mardia, 1998; Zelditch et al., 2004).

Resistant-Fit Methods.—Conventional Procrustes methods (above) rotate configurations by minimizing the square root of the summed squared distance between landmarks, and assume that variance is equal and circularly random at each landmark. However, biological differences in shape may actually be localized at one or a few landmarks (the "Pinocchio effect"). In such circumstances, conventional Procrustes superimposition

methods would result in transfer of variance from the more variable landmarks to the less variable landmarks, giving a misleading impression of the structure of shape variation. Resistant-fit methods rotate configurations by more robust optimization criteria that are less sensitive to large displacements at few landmarks.

There are many robust optimization criteria (see Press et al., 1988). The most commonly used is resistant-fit theta rho analysis (RFTRA), which uses "repeated medians" as the optimization for rescaling and rotation (Siegel and Benson, 1982; Rohlf, 1990; Chapman, 1990; Zelditch et al., 2004, pp. 119-122). Although RFTRA is less sensitive to the Pinocchio effect, it has several disadvantages that have limited its broad appeal. Rotation during the superimposition does not minimize the partial Procrustes distance between the target and reference configurations, so the configurations do not lie in shape space and do not differ solely in shape. Like conventional Procrustes methods, RFTRA permits variance at all landmarks (resulting in $2k$ free variable coordinates) but the configurations have $2k - 4$ dimensions degrees of freedom. Standard statistical tests cannot be applied to determine significance of difference in shape, although resampling methods might be applicable. No baseline is specified, so the orientation of biologically relevant axes is not respected during rotation.

BASELINE SUPERIMPOSITION METHODS

Baseline superimposition methods require designation of a baseline, i.e., an axis defined by two landmarks, one at each end. Configurations are then translated and rotated to a common baseline orientation. Depending the baseline method employed, either the x-, the y-, and/or the z- coordinate of the baseline-defining landmarks will have zero variance

following the superimposition. The baseline may also be the only factor used to rescale the configuration. Details of two baseline methods are provided below.

Bookstein (or Two-Point) Registration.—This is the most widely used baseline superimposition method. Two landmarks are selected to act as endpoints of a baseline that will be used to place all configurations into a common alignment and scale. Configurations are translated, rotated, and rescaled such that one baseline-defining landmark ends up with the x-, y-coordinates [0, 0] and the other with the x-, y-coordinates [1, 0] (below). The baseline then defines a new coordinate system within which all other landmarks can be considered (Bookstein, 1984, 1986, 1991). Landmark coordinates within this new system are often called Bookstein coordinates (BC). After superimposition by Bookstein registration, any difference in shape between configurations is summarized by offset in the position of the free (i.e., non-baseline) landmark(s) (Fig. 3.2 in main paper).

The mathematical procedure for conducting Bookstein registration is given here. The landmarks at ends of the baseline we will refer to as A and B, which are typically fixed with A at the Cartesian coordinates [0, 0] and B at coordinates [1, 0] (Bookstein, 1984, 1986, 1991). (Some authors fix these landmarks at coordinates [-0.5, 0] and [0.5, 0] [see Dryden and Mardia, 1998].)

Translation: Translation of the landmark configuration is achieved by subtracting the x- and y-coordinate values of baseline landmark A from the x- and y-coordinate values of each other landmark *i*.

New x-coordinate of landmark *i*: $x'_i = x_i - x_A$

New y-coordinate of landmark *i*: $y'_i = y_i - y_A$

where x and y are the x- and y-coordinates, respectively, of landmark *i* or A (denoted by subscript).

Rotation: Rotation is achieved by the standard matrix rotation equation for a rotation of -θ:

New x-coordinate of landmark *i*: $x''_i = x'_i\cos\theta + y'_i\sin\theta$

New y-coordinate of landmark *i*: $y''_i = y'_i\cos\theta - x'_i\sin\theta$

where $\cos\theta = (x_B - x_A)/d$ and $\sin\theta = (y_B - y_A)/d$, d is the original baseline length (calculated through trigonometry), and the subscript denotes landmark *i*, A, or B.

Rescaling: All configurations are rescaled to unit baseline length by dividing the x- and the y-coordinate values of each landmark by the original baseline length:

New x-coordinate of landmark *i*: $x'''_i = x''_i/d$

New y-coordinate of landmark *i*: $y'''_i = y''_i/d$

These three simple operations can be combined into a single formula, so that the Bookstein x-coordinate of any landmark *i* (BCx_i) is found as:

$$BCx_i = \frac{(x_B - x_A)(x_i - x_A) + (y_B - y_A)(y_i - y_A)}{(x_B - x_A)^2 + (y_B - y_A)^2}$$

and the Bookstein y-coordinate of any landmark *i* (BCy_i) is found as:

$$BCy_i = \frac{(x_B - x_A)(y_i - y_A) - (y_B - y_A)(x_i - x_A)}{(x_B - x_A)^2 + (y_B - y_A)^2}$$

where subscripts denote the x- or y-coordinate values of landmarks A, B, or *i* as digitized.

In terms of the criteria for superimposition, Bookstein registration differs from partial Procrustes superimposition in two important ways. First, configurations are

rescaled to unit baseline length rather than unit (or even the same) centroid size, meaning that size differences between the configurations are not entirely removed (because non-baseline landmarks contribute to centroid size but not baseline length; see above). Second, configurations are rotated to common baseline orientation rather than to minimize partial Procrustes distance, meaning that rotational differences between the configurations are not entirely removed (because offset between configurations in the location of corresponding non-baseline landmarks is not taken into account). Because neither size nor rotational differences between configurations is minimized by Bookstein registration, configurations superimposed using this method do not differ solely in shape (i.e., Bookstein registration does not place configurations in shape space [below]).

The results of superimposition using Bookstein registration can be interpreted and described in terms of landmark displacement relative to the invariant baseline. This is especially useful if the baseline is an important biological axis (e.g., body axis, line of symmetry) and, unlike for Procrustes superimposition methods (above), the orientation of the baseline is respected (compare Fig. 3.1 to Fig. 3.2 in the main paper). However, Bookstein registration has several disadvantages. The perception of the results depends entirely on the choice of baseline: perceived variance in non-baseline landmarks will change if a different baseline is selected (although this may not cripple attempts to investigate differences in shape; Bookstein, 1991, pp. 130-133). Furthermore, because the baseline endpoints are permitted zero variance by the superimposition, any variance in location of these landmarks is transferred to other landmarks. This transfer is biased: the variance transferred to any non-baseline depends on the distance of that landmark from the baseline. This transfer of variance can induce correlations among landmarks. The

following may be offered as a rough guide for minimizing these problems when selecting a baseline: (1) don't define a baseline using landmarks that are hard to digitize; (2) results are easier to interpret if the baseline is a meaningful body axis (i.e., one that isn't expected to show biological differences in orientation among individuals); and (3) make the baseline as long as possible (to minimize the effect of localized shape variation).

Bookstein coordinates can be used to test for differences in mean shape among groups. For k landmarks digitized in 2D there are $k - 2$ free landmarks and therefore $2k - 4$ free coordinate variables (Bookstein coordinates) following the superimposition. This equals the number of d.f. by which the configurations can differ (because one d.f. is lost to rescaling, another is lost to rotating, and two more are lost to translation). After exclusion of the fixed baseline endpoint coordinates, standard statistical tests such as Hotellings T^2 test and MANOVA can therefore be carried out to determine the significance of shape difference without discarding additional variables (note that this is not true if the configuration includes semilandmarks.) Mathematical details of these tests can be found in many introductory texts on multivariate statistics (e.g., Chatfield and Collins, 1980; Dryden and Mardia, 1998; Zelditch et al., 2004). Hotelling's T^2 test of among-group difference in mean shape based on Bookstein coordinates can be performed in the IMP program TwoGroup (Appendix 2.6; the invariant baseline endpoint coordinates are automatically discarded in this program). If more than two groups are to be compared, the Bookstein coordinates generated in CoordGen (Appendix 2.3) can be imported into most statistical software packages (e.g., Systat, SPSS, R) for analysis in a MANOVA.

However, there are several caveats of using Bookstein coordinates to test for difference in mean form between groups. Inference of difference in mean form between groups may depend on the choice of baseline (above). Bookstein registration does not entirely remove differences in coordinates of corresponding landmarks attributable to non-shape factors (above). Finally, the statistical power of tests using Bookstein coordinates is generally weaker than that of tests based on Procrustes coordinates or warp scores derived from Procrustes coordinates (Rohlf, 2000, 2003b).

Sliding Baseline Registration.—Sliding baseline registration (SBR; Webster et al., 2001; Kim et al., 2002; Zelditch et al., 2004, pp. 109-113) also requires specification of a baseline defined by endpoint landmarks A and B. However, unlike Bookstein registration, configurations are rescaled to unit centroid size (rather than unit baseline length), and the baseline is permitted to exhibit some variance in its location in one dimension. Configurations are translated and rotated such that the x-coordinate of the centroid is fixed at [x = 0], and the y-coordinates of both baseline endpoints are fixed at [y = 0]. The x-coordinates of the baseline landmarks are therefore determined by the centroid rather than being fixed.

As for Bookstein registration, the orientation of the baseline is respected and results are easy to describe and interpret in terms of landmark displacement relative to the baseline. Like partial Procrustes superimposition, SBR removes all size differences between configurations by rescaling to unit centroid size. Transfer of variance among landmarks is also reduced relative to Bookstein registration by allowing variance in the y-coordinates of the baseline endpoints (and also in baseline length).

However, transfer of variance is not entirely removed (because the y-coordinates of the baseline endpoint landmarks are fixed) and this may induce covariance among other landmarks. The perception of variance is still dependent upon choice of baseline. SBR does not place the configurations in shape space, because differences in rotation and location are not entirely removed (rotation is determined only the baseline, and configurations are not completely centered). Furthermore, SBR fixes the y-coordinates of two landmarks, resulting in (for 2D data) $2k - 2$ free coordinate variables but $2k - 4$ degrees of freedom (one lost to rescaling; one lost to rotating; and two lost to translation). This precludes conducting standard statistical tests on SBR coordinates to determine significance of difference in shape, although nonparametric resampling methods are applicable.

WHICH SUPERIMPOSITION METHOD TO USE?

Visualization of variation in form (and the extent to which this represents true shape variation) is dependent upon choice of superimposition method. This is inconvenient but unavoidable when dealing with geometric morphometric data. The extent to which statistical inference of shape difference is affected will vary from case to case. It has been demonstrated that under most circumstances GPA produces the most robust and accurate estimate of mean shape, and that statistical approaches based on Procrustes superimposition have greater power, relative to Bookstein registration and EDMA (Rohlf, 2000, 2003b). As a general rule of thumb it is therefore advisable to conduct statistical approaches on Procrustes-based variables and to use other superimposition methods as visualization or exploratory tools. Nevertheless, the particulars of a given case study may

justify favoring a different superimposition method. The CoordGen software allows for easy and rapid comparison of results produced by the different methods, or by different baseline choices, so that the researcher can determine by comparison which features of a result are highly influenced by the choice of method and which are invariant (Appendix 2.3 in the main paper). Many of the other IMP and tps software programs allow display of data in more than one superimposition.

One method for analysis of morphometric data, EDMA, does not rely on superimposition for comparison of form and therefore avoids the above issue (Lele, 1991, 1993; Lele and Richtsmeier, 1991; Richtsmeier and Lele, 1993; Richtsmeier et al., 2002). EDMA compares forms based on all pairwise interlandmark distances in each configuration. Statistical problems with EDMA have been demonstrated (Rohlf, 2000, 2003b; Adams et al., 2004), and the method is not further discussed here, although it may prove useful in specialized applications.

QUANTIFYING THE AMOUNT OF SHAPE VARIATION

Sample Variation and Disparity.—Variation in shape within a sample can be quantified as the average partial Procrustes distance away from the mean form for the sample (based on Foote; 1993):

$$\text{Variation in shape} = \frac{\sum_{i=1 \text{ to } g}\sum_{j=1 \text{ to } n} d_{ij}^2}{N - 1}$$

where g = number of groups; n = number of specimens in group i; N = total number of specimens; and d_{ij}^2 = squared partial Procrustes distance of the j^{th} specimen in group i from the mean form of the total sample. This metric is appropriate for calculating

measurement error for a single specimen, variation within a single species (when $g = 1$, $n = N$, and the mean form of the total sample = the group mean), or disparity among species.

Partial Disparity.—Partial disparity (PD; Foote, 1993) is a metric that quantifies the contribution of a particular group (i) to the total morphological disparity of a more inclusive sample:

$$PD_i = \frac{\sum_{j = 1 \text{ to } n} d_{ij}^2}{N - 1}$$

where n = number of specimens in group i; N = total number of specimens; and d_{ij}^2 = squared partial Procrustes distance of the j^{th} specimen in group i from the mean (consensus) form of the total sample. (Note that the partial disparity of group i is *not equivalent* to within-group variation of group i [above] unless the total sample includes only group i.) It follows that the total morphological disparity within the sample is equal to the sum of partial disparities of all included groups.

SHAPE AND SHAPE SPACES

The theory of shape is the key concept of geometric morphometrics. An understanding of the theory of shape allows determination of the validity of particular morphometric methods for particular questions. Shape is all the geometric information that remains when location, scale, and rotational effects are filtered out from an object (Kendall, 1977). Translation, rescaling, and rotation do not affect the shape of a landmark configuration, but any other operation does.

There are many "spaces" within which landmark configurations can be compared. The extent to which those comparisons are meaningful (i.e., the extent to which non-shape differences between the configurations have been removed) depends on the operations that have been performed to superimpose the configurations. Each operation during a superimposition results in placing the configurations in a difference space. Four such spaces are considered in this review. Clear, introductory reviews can also be found in O'Higgins (2000), Zelditch et al. (2004, pp. 73-104), and Slice (2007). A detailed discussion of the issues involved and a comparison of the statistical power of the different methods can be found in Rohlf (2000).

Configuration Space.—This is the space defined by the raw landmark coordinates, as originally digitized and with no operations performed to superimpose configurations. Any two configurations therefore differ in both shape and non-shape parameters. For landmark configurations, the dimensionality of configuration space is km, where k is the number of landmarks and m is the number of planes in which the landmarks were digitized (2 or 3). The location of any shape in that configuration space is specified by km components (the values in the landmark data matrix), and configurations are free to differ by km degrees of freedom.

Shape Space.—Configurations that have been centered and scaled to unit centroid size are said to be in pre-shape space. When they are also optimally rotated using a least-squares fit to a given reference form (i.e., configurations that have been placed in partial Procrustes superimposition) they then lie in shape space (sometimes called "oriented pre-

shape space"). This is the space in which shapes are optimally compared, because all non-shape differences between configurations have been removed, but centroid size is fixed at 1. Configurations in shape space are free to differ by $km - m - 1 - \{m(m - 1)/2\}$ degrees of freedom, because the process of translation results in loss of m d.f., the process of rescaling results in loss of 1 d.f., and the process of rotation results in loss of another $\{m(m - 1)/2\}$. Shape space therefore has $2k - 4$ dimensions for 2D data, and $3k - 7$ for 3D data.

The nature of shape space is easiest to comprehend when dealing with landmark configurations of triangles. Configuration space for triangles has $km = 6$ dimensions. As a result of placing triangles in partial Procrustes superimposition, two dimensions are lost by translating to common centroid location, one dimension is lost by rescaling to unit centroid size, and one more dimension is lost by rotating. Shape space for triangles is therefore two-dimensional (a surface), and any two triangles in shape space are separated by two degrees of freedom. Since shape space has no "edges", the two-dimensional surface can be envisioned as the (2-D) surface of a (3-D) hollow ball (or hemisphere). The radius of the ball is 1 (because configurations were rescaled to unit centroid size), and the ball is centered on coordinates [0, 0, 0]. Every possible configuration of three landmarks (in partial Procrustes superimposition with a given reference form) is represented as a point on the surface of the hemisphere (Kendall, 1984; Goodall, 1991; Small, 1996). If the reference form lies at the pole, then "lines of latitude" on the surface represent shapes of equal (partial) Procrustes distance from the reference form (below). The location of each shape relative to the reference form is given by two parameters (the degrees of freedom).

Generalizing beyond triangles, shape space for any form is a [km - m - 1 - {m(m - 1)/2}]-dimensional surface of a hypersphere centered on the origin and with a radius of one (Kendall, 1984; Goodall, 1991; Small, 1996). All shapes (in partial Procrustes superimposition) are represented by points on that surface, and the location of any shape relative to another in shape space is specified by km - m - 1 - {m(m - 1)/2} components (the degrees of freedom).

Quantifying the Amount of Difference in Shape Between Configurations.—The amount of difference in shape between any two configurations is represented by the magnitude of their separation on the surface of shape space. This is most intuitively quantified as the distance over the surface of the shape space hypersphere between the points representing those configurations. This metric is known as the *Procrustes distance* (ρ) between the configurations. Because the radius of the hypersphere is one, the Procrustes distance ρ between any two forms is equal to the angle (in radians) between the radii that connect the center of the hypersphere to their respective locations on the surface. Values for ρ range from 0 to $\pi/2$.

However, the shortest distance between the two points passes through the inside of the hollow hypersphere, not over the surface. This chord is the *partial Procrustes distance* (D_p), and has been encountered (main text; above) as the square root of the summed squared distances between homologous landmarks on two configurations following partial Procrustes superimposition. At small distances, the chord length is a reasonable approximation of the arc length.

The distance separating the two configurations can be further reduced if the centroid sizes of those configurations are not constrained to unity. The distance is truly minimized when the centroid size of one configuration (the reference) is 1 and that of the other is reduced to cos(ρ). Under these conditions, the metric is the *full Procrustes distance* (D_F) between those configurations. Fortunately, the three measures of distance between forms are related:

$$D_p = \text{chord of arc } \rho = 2\sin(\rho/2)$$

$$D_F = \sin(\rho)$$

Procrustes distance is rarely reported because it is an arc-length in the curved, non-Euclidean shape space and very specialized methods are required for statistical inference. Of the two linear measures, partial Procrustes distance is preferentially reported over full Procrustes distance because it more closely approximates Procrustes distance (Dryden and Mardia, 1998; Zelditch et al., 2004).

Kendall's Shape Space.—The configurations of all shapes can be rescaled to their minimal distance (D_F) from a given reference form. When the centroid size of reference form is set to 1 and the centroid size of every other form = cos(ρ) (where ρ is the Procrustes distance from reference form), the shapes define a new hyperspherical shape space of radius 0.5, nestled within and tangent to shape space at the reference form. This is Kendall's shape space (Kendall, 1984). Configurations within Kendall's shape space are in *full Procrustes superimposition* with respect to the reference form. Kendall's shape space has the same dimensionality as shape space (rescaling results in loss of 1 degree of

freedom irrespective of whether configurations are rescaled to unit centroid size [shape space] or to cos(ρ) [Kendall's shape space]).

Tangent Space.—Shape space and Kendall's shape space are curved (non-linear, non-Euclidean) spaces. However, most statistical tools assume a linear, Euclidean space. It is therefore typical to use a Euclidean space as a proxy for (Kendall's) shape space in order to apply statistical tests to shape data. The Euclidean space is tangent to (Kendall's) shape space, and is commonly referred to as the tangent space. Operations such as conducting a PCA of landmark coordinates (or on the uniform component and partial warp scores as discussed below) are most easily carried out in Euclidean tangent plane: the tangent point is then the reference form to which all other configurations were aligned using Procrustes superimposition.

Working in any tangent space comes at a cost. The Euclidean space cannot exactly replicate the distances between configurations in shape space. Rather, distances of shapes away from the reference form will be distorted. This is analogous to the distortion of distances between geographic locations when a cartographer projects from a globe onto a map surface. The further the distance from the reference form, the more severe the distortion. Distortion of distances from target configurations to the reference configuration resulting from this projection is globally minimized when the reference form lies centrally within the range of shapes. For this reason, the default procedure is to use the mean (consensus) form of all configurations in the sample as the reference form. The tpsSmall program (Rohlf, 2003a) can provide information about the distortion caused by using the tangent space approximation (below), and its use should be considered if

large partial Procrustes distances (over 0.1 as a rule of thumb) are common. However, it is sometimes desirable to use a non-central shape as reference form. This might arise when shape deviation from a specific form (e.g., an ancestor or an early ontogenetic stage) is of interest. Average distortion when projecting from shape space into tangent space will be larger in such situations.

Checking for distortion when projecting from shape space to tangent space.—It is critical to determine whether the distortion introduced by this projection is of a magnitude sufficient to influence analytical results (see text). This can be achieved by (1) calculating the Procrustes distance between two forms and the Euclidean distance between those same forms in tangent space; then (2) calculating the correlation between these variables, assessed over all pairwise comparisons in the sample (Rohlf, 1999). Such a regression can be conducted in the program tpsSmall (Rohlf, 1998, 2003a; available through the SUNYSB site).

A landmark data file in *TPS* format (e.g., generated by tpsDig, or converted into this format using CoordGen) can be loaded into tpsSmall by clicking on the "Data" push-button. (Another format, the "*NTS*", is also acceptable; see the tpsSmall Help file for details.) The analysis is performed by clicking the "Compute" push-button. The program computes a reference form (the mean of all forms in the sample) and superimposes all configurations onto that reference using GPA. For all pairwise comparisons of configurations in the sample, the program calculates both the Procrustes distance and the Euclidean distance between forms (based on superimposition to the reference form, which is central in the distribution). The regression coefficients of the correlation

between these variables are given in a report window. A bivariate plot showing the regression can be obtained by clicking the "View Plot" push-button. Options for viewing all pairwise distances or just the distances of each configuration from the reference form are available in the "Plot Options" window. (The former is more informative.)

Distortion makes the regression deviate from linearity, lowering the correlation coefficient and giving the regression a slope of less than 1. Unfortunately the non-independence of points in the regression invalidates any statistical significance test. However, even a slope of 0.96 may indicate serious distortion (noted by an anonymous reviewer). Such distortion is a result of including grossly different shapes in a sample. This is typically not encountered in biological data, although they can be produced by errors during data collection or processing (e.g., associated with inconsistent data reflection).

The "Options" pull-down menu provides the ability to perform the regression using Full Procrustes superimposition (above) and/or using alternative methods of projection into tangent space (see Rohlf [1999] for a discussion). The distortion introduced when using these alternatives will be similar to that introduced when using the default settings (partial Procrustes superimposition and orthogonal projection) unless shape variation in the sample is very large.

As an example of application of this technique to real data, all pairwise comparisons of inter-specimen distance in shape space (partial Procrustes distance) versus Euclidean distance in tangent space (using an orthogonal projection) were calculated for the total sample of *Olenellus gilberti* (n = 58), *O. chiefensis* (n = 63), *O. terminatus* (n = 37) and *O. fowleri* (n = 23) analyzed in the main paper. The slope of the

fitted regression line (partial Procrustes distance to Euclidean distance) is 0.998, and the correlation between the two distance measures is very strong (r = 0.999). Distortion of inter-specimen distances associated with the projection from shape space into a tangent space approximation of that space is considered negligible.

THE THIN-PLATE SPLINE

The thin-plate spline (TPS; Bookstein, 1989, 1991, 1996) is an interpolation function depicting how the shape of one form (the reference form) can be deformed to match the shape of a second form (the target form). The function can be used to produce deformation grids that depict smooth changes across a biological structure. The offset in location of corresponding landmarks between the two configurations provides the only data, and the TPS interpolates (using a curve-smoothing spline function) deformation in regions between the landmarks. In addition to this graphic use, the method can also be used to form a useful orthogonal basis set (via an eigenvector decomposition) whose degrees of freedom match the number of axes. A simplistic summary of the generation of the spline was given in the main paper; the mathematical details are provided here.

Initial Data.—The configuration of the target form is aligned with that of the reference form in a partial Procrustes superimposition. The configurations now differ only in shape.

Let $\mathbf{X}_{\mathbf{r}\ (k \times 2)}$ be the matrix of landmark coordinates of the reference form in this alignment, and let $\mathbf{X}_{\mathbf{t}\ (k \times 2)}$ be the matrix of landmark coordinates of the target form in this alignment.

$$\text{Landmark coordinates of reference form, } \mathbf{X}_r = \begin{pmatrix} x_{r1} & y_{r1} \\ x_{r2} & y_{r2} \\ x_{r3} & y_{r3} \\ \cdots & \cdots \\ x_{rk} & y_{rk} \end{pmatrix}$$

$$\text{Landmark coordinates of target form, } \mathbf{X}_t = \begin{pmatrix} x_{t1} & y_{t1} \\ x_{t2} & y_{t2} \\ x_{t3} & y_{t3} \\ \cdots & \cdots \\ x_{tk} & y_{tk} \end{pmatrix}$$

where x and y are the x- and y-coordinates, respectively, of landmarks 1 through k on the reference (subscript r) or target form (subscript t).

*Matrix **V**: Deviations Between Forms.*—Calculate the deviation in the x- and y-direction at each landmark between the reference and target form. Let Δx_i, Δy_i be the offset in location between landmark i on the reference form and the target form in this alignment.

$$\begin{aligned} \Delta x_i &= x_{ri} - x_{ti} \\ \Delta y_i &= y_{ri} - y_{ti} \end{aligned}$$

Construct a matrix $\mathbf{V}_{([k+3] \times 2)}$ of these deviations:

$$\text{Deviations between reference and target forms, } \mathbf{V} = \begin{pmatrix} \Delta x_1 & \Delta y_1 \\ \Delta x_2 & \Delta y_2 \\ \Delta x_3 & \Delta y_3 \\ \cdots & \cdots \\ \Delta x_k & \Delta y_k \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

The three rows of zeros added serve to constrain the weightings (see below) of the spline function fitted to these displacements, by requiring that the sum of the x- and y- weights be zero and that the sum of cross-products of these weights and the displacements also be zero.

Styles of Deformation.—The styles of deformation that are possible for a given reference configuration fall into two categories. *Uniform (affine) transformations* require no bending energy: grid lines on the spine surface that were parallel prior to the deformation remain parallel after the deformation, meaning that shape difference is uniform across the configuration. *Non-uniform (non-affine) transformations* require bending energy: grid lines on the spine surface that were parallel prior to the deformation are bent into non-parallel configurations after the deformation, meaning that shape difference is not uniform across the configuration.

Uniform Transformations.—For two dimensional data there are six mutually independent ways to cause offset in landmark location between forms without bending the deformation grid lines. Four (translation along the *x*-axis; translation along the *y*-axis; rescaling; and rotation) do not alter shape and are termed *implicit* uniform deformations. Two (compression/dilation; and shear) do alter shape and are termed *explicit* uniform deformations.

For any point on the surface, uniform deformation in the *x*-direction is summarized by three components:

a_{x1} = Uniform translation in *x*.

$a_{xx}x_r$ = Uniform deformation in *x* as a function of *x* on the reference form.

$a_{xy}y_r$ = Uniform deformation in *x* as a function of *y* on the reference form.

The second and third components together summarize deformation in the *x*-direction due to rotation, rescaling, compression/dilation, and shearing. Similarly, for any point on the

spline surface, uniform deformation in the *y*-direction is summarized by three components:

a_{y1} = Uniform translation in *y*.

$a_{yx}x_r$ = Uniform deformation in *y* as a function of *x* on the reference form.

$a_{yy}y_r$ = Uniform deformation in *y* as a function of *y* on the reference form.

Again, the second and third components together summarize deformation in the *y*-direction due to rotation, rescaling, compression/dilation, and shearing.

Non-Uniform Transformations.—These deformations involve bending (flexing, warping) of the spline surface relative to the reference surface, so that initially parallel lines are bent by the deformation. "Bending energy" is required, inversely proportional to the spacing between landmarks on the reference form. The TPS model is based on the analogy of the bending of a thin metal plate. Deformations of large spatial scale (i.e., involving differential displacement of landmarks that are widely separated on the reference configuration) require less bending energy than do deformations of a smaller spatial scale (i.e., involving differential displacement of landmarks that are more closely spaced on the reference configuration).

The spatial scale over which a given non-uniform transformation must be achieved is quantified as a *U*-function of interlandmark distances on the reference form. Let r_{ij} be the distance between landmarks *i* and *j* on the reference form:

$$\text{Distance between landmarks } i \text{ and } j \text{ on reference form, } r_{ii} = \sqrt{((x_{ri} - x_{ri})^2 + (y_{ri} - y_{ri})^2)}$$

Define a U-function for each interlandmark distance r_{ij} on the reference form:

$$U(r_{ij}) = U_{ij} = r_{ij}^2 \log (r_{ij}^2)$$

Construct the matrix $\mathbf{P}_{k\ (k\ x\ k)}$ of interlandmark distance U-functions:

$$\text{Matrix of interlandmark distance functions, } \mathbf{P}_k = \begin{pmatrix} 0 & U_{12} & \cdots & U_{1k} \\ U_{21} & 0 & \cdots & U_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ U_{k1} & U_{k2} & \cdots & 0 \end{pmatrix}$$

For any point, the total non-uniform deformation in the *x*-direction is summarized by *k U*-functions, each multiplied by a coefficient (or weight) *w*:

$w_{xi}U_{ij}$ = Non-uniform deformation in *x* relating to the *U*-function of the distance between landmarks *i* and *j* on the reference form.

Similarly, for any point, non-uniform deformation in the *y*-direction is summarized by *k U*-functions, each multiplied by a coefficient *w*:

$w_{yi}U_{ij}$ = Non-uniform deformation in *y* relating to the *U*-function of the distance between landmarks *i* and *j* on the reference form.

Drawing the Spine.—The observed deformation (**V**) is a function of the uniform deformation (*a* terms) acting on the reference configuration ($\mathbf{X}_r$), and of the non-uniform deformation (*U*-functions of the interlandmark distances on the reference form [$\mathbf{P}_k$], multiplied by their coefficients [*w* terms]). This can be represented as a linear equation:

$$\begin{aligned} \Delta x &= a_{x1} + a_{xx}x_r + a_{xy}y_r + \sum (w_{xi}U_{ij}) \\ \Delta y &= a_{y1} + a_{yx}x_r + a_{yy}y_r + \sum (w_{yi}U_{ij}) \end{aligned}$$

Note that this is a simple reorganization of the data: no information is lost when shape deviation is expressed this way. However, the coefficients of non-uniform deformations (*w*) and the uniform terms (*a*) are unknown. Matrix algebra can be used to solve for these unknowns.

Let **W** be the matrix of uniform terms and weights by which the *U*-functions must be multiplied to yield the spline surface. **W** has the dimensionality ([k + 3] x 2), where the first k rows are coefficients of the non-uniform transformations and the last three rows are the uniform transformations.

$$\text{Matrix of transformation coefficients, } \mathbf{W} = \begin{pmatrix} w_{x1} & w_{y1} \\ w_{x2} & w_{y2} \\ w_{x3} & w_{y3} \\ \dots & \dots \\ w_{xk} & w_{yk} \\ a_{x1} & a_{y1} \\ a_{xx} & a_{yx} \\ a_{xy} & a_{yy} \end{pmatrix}$$

Let **Q** be the matrix of landmark coordinates of the reference form ($\mathbf{X}_r$) with a column of 1's added. This additional column is needed to accommodate the translations (see below).

$$\text{Matrix of landmark coordinates of reference form, } \mathbf{Q} = \begin{pmatrix} 1 & x_{r1} & y_{r1} \\ 1 & x_{r2} & y_{r2} \\ 1 & x_{r3} & y_{r3} \\ 1 & \dots & \dots \\ 1 & x_{rk} & y_{rk} \end{pmatrix}$$

Construct matrix $\mathbf{L}_{([k+3] \text{ x } [k+3])}$, containing the landmark coordinates and the *U*-functions of the interlandmark distances of the reference form.

$$\text{Matrix } \mathbf{L} = \begin{pmatrix} \mathbf{P}_k & \mathbf{Q} \\ & 0\ 0\ 0 \\ \mathbf{Q}^T & 0\ 0\ 0 \\ & 0\ 0\ 0 \end{pmatrix}$$

$$\text{Matrix } \mathbf{L} = \begin{pmatrix} 0 & U_{12} & \dots & U_{1k} & 1 & x_{r1} & y_{r1} \\ U_{21} & 0 & \dots & U_{2k} & 1 & x_{r2} & y_{r2} \\ \dots & \dots & \dots & \dots & 1 & \dots & \dots \\ U_{k1} & U_{k2} & \dots & 0 & 1 & x_{rk} & y_{rk} \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ x_{r1} & x_{r2} & \dots & x_{rk} & 0 & 0 & 0 \\ y_{r1} & y_{r2} & \dots & y_{rk} & 0 & 0 & 0 \end{pmatrix}$$

We can now summarize the deformation in matrix form.

$$\mathbf{LW} = \mathbf{V}$$

$$\begin{pmatrix} 0 & U_{12} & \dots & U_{1k} & 1 & x_{r1} & y_{r1} \\ U_{21} & 0 & \dots & U_{2k} & 1 & x_{r2} & y_{r2} \\ \dots & \dots & \dots & \dots & 1 & \dots & \dots \\ U_{k1} & U_{k2} & \dots & 0 & 1 & x_{rk} & y_{rk} \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ x_{r1} & x_{r2} & \dots & x_{rk} & 0 & 0 & 0 \\ y_{r1} & y_{r2} & \dots & y_{rk} & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} w_{x1} & w_{y1} \\ w_{x2} & w_{y2} \\ \dots & \dots \\ w_{xk} & w_{yk} \\ a_{x1} & a_{y1} \\ a_{xx} & a_{yx} \\ a_{xy} & a_{yy} \end{pmatrix} = \begin{pmatrix} \Delta x_1 & \Delta y_1 \\ \Delta x_2 & \Delta y_2 \\ \dots & \dots \\ \Delta x_k & \Delta y_k \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

The linear equations of this matrix operation define the deformation of the thin-plate spline surface (as above, but written as discrete linear equations):

Linear equations for x-deviations:

$$\begin{aligned} w_{x1}U_{11} + w_{x2}U_{12} + \dots + w_{xk}U_{1k} + a_{x1} + a_{xx}x_{r1} + a_{xy}y_{r1} &= \Delta x_1 \\ w_{x1}U_{21} + w_{x2}U_{22} + \dots + w_{xk}U_{2k} + a_{x1} + a_{xx}x_{r2} + a_{xy}y_{r2} &= \Delta x_2 \\ \dots & \\ w_{x1}U_{k1} + w_{x2}U_{k2} + \dots + w_{xk}U_{kk} + a_{x1} + a_{xx}x_{rk} + a_{xy}y_{rk} &= \Delta x_k \\ w_{x1} + w_{x2} + \dots + w_{xk} + 0 + 0 + 0 &= 0 \\ w_{x1}x_{r1} + w_{x2}x_{r2} + \dots + w_{xk}x_{rk} + 0 + 0 + 0 &= 0 \\ w_{x1}y_{r1} + w_{x2}y_{r2} + \dots + w_{xk}y_{rk} + 0 + 0 + 0 &= 0 \end{aligned}$$

Linear equations for y-deviations:

$$\begin{aligned} w_{y1}U_{11} + w_{y2}U_{12} + \dots + w_{yk}U_{1k} + a_{y1} + a_{yx}x_{r1} + a_{yy}y_{r1} &= \Delta y_1 \\ w_{y1}U_{21} + w_{y2}U_{22} + \dots + w_{yk}U_{2k} + a_{y1} + a_{yx}x_{r2} + a_{yy}y_{r2} &= \Delta y_2 \\ \dots & \\ w_{y1}U_{k1} + w_{y2}U_{k2} + \dots + w_{yk}U_{kk} + a_{y1} + a_{yx}x_{rk} + a_{yy}y_{rk} &= \Delta y_k \\ w_{y1} + w_{y2} + \dots + w_{yk} + 0 + 0 + 0 &= 0 \\ w_{y1}x_{r1} + w_{y2}x_{r2} + \dots + w_{yk}x_{rk} + 0 + 0 + 0 &= 0 \\ w_{y1}y_{r1} + w_{y2}y_{r2} + \dots + w_{yk}y_{rk} + 0 + 0 + 0 &= 0 \end{aligned}$$

The unknown coefficients in the linear equations (contained within matrix **W**) can now be solved by simple rearrangement after the computation of the inverse of L.

$$\mathbf{LW} = \mathbf{V}$$

$$\mathbf{W} = \mathbf{L}^{-1}\mathbf{V}$$

All terms in the linear equation are now known, and the deformation in the *x*-direction and the deformation in the *y*-direction can be calculated for any point on the reference surface:

$$\begin{aligned} \Delta x &= a_{x1} + a_{xx}x_r + a_{xy}y_r + \sum (w_{xi}U_{ij}) \\ \Delta y &= a_{y1} + a_{yx}x_r + a_{yy}y_r + \sum (w_{yi}U_{ij}) \end{aligned}$$

This is a continuously differentiable function in which the second derivative has been minimized to minimize bending energy. This particular form of the TPS is particularly useful when programming routines to plot the interpolated deformations, due to its ease of use.

WARP SCORES

Warp scores are calculated through mathematical decomposition of the bending energy matrix associated with the thin-plate spline, followed by calculation of the component (or loading) of each specimen along each eigenvector (or principal warp) of the decomposition. This is an approach very similar to principal component analysis (PCA). In a PCA, the eigenvector analysis is done on the variance-covariance matrix, so that the PC axes derived organize information based on the variance explained by the axes. In a principal warp computation, information is ordered by spatial scale. Each warp score quantifies the contribution of a mathematically independent style of deformation (warp) to the shape difference between the reference and target forms. The mathematical derivation of warps (principal warps) and warp (or partial warp) scores is presented below.

There are two kinds of warp scores: partial warp scores (relating to non-uniform deformation), and uniform warp scores (also called the explicit uniform terms).

Partial Warps.—Partial warps are often described as "idealized modes of bending". A partial warp defines non-uniform deformation of the reference form at a particular spatial scale. Deformation at that spatial scale is mathematically independent of deformation at all other spatial scales. Considered together, the partial warps fully describe mathematically independent modes of non-uniform deformation of the reference form.

Partial warps are a function of the reference form alone and have no magnitude or direction. A magnitude and a direction for each partial warp are imparted when the reference form is compared to a target form. These are expressed in the *partial warp scores* for that partial warp in that comparison.

Deriving Partial Warps.—Construct a new matrix $\mathbf{L}_k^{-1}$, consisting of the upper left [k x k] block of matrix $\mathbf{L}^{-1}$ (see details of thin-plate spline, above). This is the "bending energy matrix". Note that only data from the landmark configuration of the reference form are involved in the construction of matrix **L**: the bending energy matrix is therefore a function only of the reference form.

Conduct an eigenanalysis of the bending energy matrix $\mathbf{L}_k^{-1}$ (dimensionality k x k), just is as done in a PCA. Due to the non-independence of interlandmark distances, three of the eigenvalues will be zero. There will therefore be $k - 3$ non-zero eigenvalues and $k - 3$ relevant eigenvectors, called the *principal warps* (each of dimensionality k x 1).

Let **E** be the matrix of eigenvectors of non-zero eigenvalues of the bending energy matrix. This matrix has a dimensionality of k x (k – 3). The non-zero eigenvectors of $\mathbf{L}_k^{-1}$ define a set of derived axes in a new morphospace. Each derived axis represents a mode of deformation at a particular spatial scale on the reference form. (Remember that matrix $\mathbf{P}_k$, which formed the analogous part of matrix **L**, was constructed from the U-functions of interlandmark distances on the reference form.) Each derived axis is a partial warp, and is orthogonal to (and therefore mathematically independent of) all others. The derived axes together define a particular basis of the Euclidean (linear) space, tangent to shape space at the reference form. The tangent morphospace is defined by axes representing modes of deformation of the reference form at mathematically independent spatial scales, which are typically ordered by spatial scale. Note that no information is lost in transforming data from Cartesian coordinates to partial warp scores: rather it is simply reorganized into a form that is mathematically convenient for many problems. It is more convenient to use these scores for statistical analysis (rather than working directly with the coefficients discussed earlier) due to the ease of depicting the results and the degree of freedom issues.

Each eigenvalue of $\mathbf{L}_k^{-1}$ is the bending energy required to achieve one unit of shape change (measured in Procrustes distance units) at that particular spatial scale. Bending energy is inversely proportional to the spatial scale of deformation: one unit of shape change is relatively easy to achieve over a large spatial scale but relatively hard to achieve over a small spatial scale. Ordering the partial warps by eigenvalue (small to large) is equivalent to ordering modes of deformation of the reference form by progressively smaller spatial scale.

Calculating Partial Warp Scores.—A partial warp score represents the actual contribution of a particular partial warp to a particular shape comparison. Each partial warp score is expressed as two components, one representing the magnitude of deformation in the *x*-direction along the axis of shape change described by the partial warp, and the other representing the magnitude of deformation in the *y*-direction along the axis of shape change described by the partial warp. Together, the two components of a partial warp score define the magnitude and direction of deformation along that partial warp axis for that particular shape comparison. For 2-D data there are $k - 3$ partial warps, and therefore $2k - 6$ partial warp score components.

To calculate the partial warp scores, first find the deviation in the *x*-direction at landmark *i* between the reference form and target form. Construct a column vector $\mathbf{v}_x$ of *x*-deviations for all *k* landmarks (this is equal to the first *k* rows in the first column of matrix **V**, above). Repeat this for deviations in the *y*-direction to construct column vector $\mathbf{v}_y$ (equal to the first *k* rows in the second column of matrix **V**, above). The *x*-components of the partial warp scores are given by $\mathbf{v}_x^T\mathbf{E}$, where **E** is the (k x [$k - 3$]) matrix of eigenvectors of the bending energy matrix. The *y*-components of the partial warp scores are given by $\mathbf{v}_y^T\mathbf{E}$.

Uniform Warps.—Six uniform components were defined in calculation of the function describing the thin-plate spline surface (above): a_{x1} and a_{y1} (uniform translation in *x* and in *y*); $a_{xx}x_r$ and $a_{yx}x_r$ (uniform deformation in *x* and in *y* as a function of *x* on the reference form); and $a_{xy}y_r$ and $a_{yy}y_r$ (uniform deformation in *x* and in *y* as a function of *y* on the

reference form). The last four components together summarize deformation in the *x*-direction and in the *y*-direction due to rotation, rescaling, compression/dilation, and shearing.

However, the reference and target forms were placed in partial Procrustes superimposition prior to calculation of their shape differences. All non-shape differences (translation in *x*, translation in *y*, size, and rotation) had therefore already been removed. The configurations could only differ in the explicit uniform components (compression/dilation and shear), which alter shape. Only two (not six) terms are actually needed to describe the remaining uniform components of shape difference. These terms describe the contribution of compression/dilation and of shearing to the total deformation. Unfortunately, the *a* terms above cannot readily be reduced to two terms: they are an overspecification of the deformation.

Calculating the Uniform Terms.—The currently favored formulae for calculating the uniform compression/dilation and shear terms were presented by Bookstein (1996; see also Zelditch et al., 2004, pp. 137-142), although alternative methods have been proposed. The configuration of the reference form is in shape space, with a centroid size of 1 and with coordinates centered on the centroid. It should also be rotated into principal axes alignment, such that $\sum_j x_j y_j = 0$, where j = landmarks 1 through k. Note that after the calculations have been completed, both the reference and all target forms may be rotated to any desired angle without changing any of the statistical results, as rotation is not part of shape (Kendall, 1977). The requirement that the reference be rotated to principal axes

orientation is a computational convenience only, not a fundamental element of the method.

The vector $\mathbf{v}_1$ describing the idealized shear component of uniform deformation of the reference form (of unit length, and projecting the sheared form back into shape space) is given by:

$$\mathbf{v}_1 = \left\{ \left(y_j \sqrt{(\alpha / \gamma)} , x_j \sqrt{(\gamma / \alpha)} \right) \right\}_{j=1 \text{ to } k}$$

where: $\alpha = \sum_i x_j^2 \qquad \gamma = \sum_i y_j^2 \qquad \alpha + \gamma = 1$

The vector $\mathbf{v}_2$ describing the idealized compression/dilation component of uniform deformation of the reference form (at unit length, and projecting the compressed/dilated form back into shape space) is given by:

$$\mathbf{v}_2 = \left\{ \left(- x_j \sqrt{(\gamma / \alpha)} , y_j \sqrt{(\alpha / \gamma)} \right) \right\}_{j=1 \text{ to } k}$$

Vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are functions of the reference form alone. The magnitude and direction of these vectors depend on the target form.

The uniform deformation resulting from shear in any given comparison is given by the summed cross-products of the deviations between forms with the shear vector $\mathbf{v}_1$.

$$\text{Shear term} = \frac{(\alpha \sum y_j \Delta x_j + \gamma \sum x_j \Delta y_j)}{\sqrt{(\alpha \gamma)}}$$

The uniform deformation resulting from compression/dilation in any given comparison is given by the summed cross-products of the deviations between forms with the compression/dilation vector $\mathbf{v}_2$.

$$\text{Compression/dilation term} = \frac{(-\gamma\sum x_j\Delta y_j + \alpha\sum y_j\Delta x_j)}{\sqrt{(\alpha\gamma)}}$$

Describing Shape Difference.—The total difference in shape between the reference form and a target form is therefore described by $2k - 4$ parameters: $2k - 6$ non-uniform terms (the partial warp scores in the *x*- and in the *y*-direction for each of the $k - 3$ partial warps), and 2 uniform terms (the difference in shape resulting from uniform compression/dilation and shear). The number of free parameters (non-uniform plus uniform terms) now equals the number of degrees of freedom by which the configurations differ in shape space (unless the configuration included semilandmarks), and standard statistical analyses (e.g., Hotelling's T^2) can be conducted on the full suite of warp scores to test for the significance of any difference in shape between groups.

Although partial warps and uniform terms of any given shape transformation relate to mathematically independent styles of deformation of the reference form, they cannot be treated as biologically independent and should not be analyzed or interpreted in isolation.

STATISTICAL TESTING OF GROUP ASSIGNMENTS

Canonical variates analysis (CVA; Text Box 4) can be used to determining whether the a priori group assignments of specimens is statistically supported by shape variables. (Strong support is not necessarily expected: the a priori classification need not have been based on shape.) It can also be used to determine the best group membership of an

unclassified specimen (assuming that the specimen actually belongs in one of the pre-defined groups).

One approach is Fisher's linear discriminant function (see Mardia et al., 1979). In this approach one seeks vectors that maximize the ratio of variance between the observed groups relative to the variance within the observed groups. This is done by carrying out an eigenvector analysis of the inverse of the within-group variance-covariance matrix times the between-group covariance matrix. The resulting eigenvectors (the CV axes) can then be tested to determine how many of them are statistically distinct (i.e., meaningful) using a Wilk's λ test. The CV axes are typically scaled so that the within-group variance along the axis is scaled to 1. Scores for each specimen are then obtained along each significant CV axis.

Statistical group assignment is based on the distance (D) between the specimen and each of the group means:

$$\text{distance, D} = \sqrt{(\mathbf{d}^T \mathbf{S}_w^{-1} \mathbf{d})}$$

where matrix $\mathbf{S}_w$ is the weighted pooled within-group variance-covariance matrix and vector **d** is the column vector of differences between specimen *x* and the mean of group *i* on all *p* variables (Mardia et al., 1979). Incorporation of matrix $\mathbf{S}_w$ ensures that within-group variance and covariance are into account. The distance used in this calculation is not a Procrustes distance, because the axes have been rescaled. CVA thus distorts distances between specimens in a way that the TPS decomposition or a PCA does not. The predicted group membership of a specimen is the group mean from which it has the lowest D. This lowest D may still be large: in such a circumstance the researcher should consider whether the specimen does not belong in any of the pre-defined groups. The

statistical confidence with which specimen *x* can be assigned to group *i* can be found by calculating the F-statistic of the distance D from *x* to group *i*:

$$F_s = \frac{1 + n_i - p - 1}{(1 + n_i - 2)p} \cdot \frac{n_i}{1 + n_i} \cdot D^2$$

where n_i is the sample size in group *i* and *p* is the number of variables, and comparing this F-statistic to critical values on an F-table (with *p* and $[1 + n_i - p - 1]$ degrees of freedom). However, this approach is thought to be rather optimistic and will probably overstate the probability (Mardia et al., 1979). A more robust approach may be to consider a *cross-validation* rate of assignment (see Text Box 4).

REFERENCES

ADAMS, D. C., F. J. ROHLF, AND D. E. SLICE. 2004. Geometric morphometrics: ten years of progress following the 'revolution'. Italian Journal of Zoology, 71:5-16.

ANDRESEN, P. R., F. L. BOOKSTEIN, K. CONRADSEN, B. ERSBØLL, J. MARSH, AND S. KREIBORG. 2000. Surface-bounded growth modeling applied to human mandibles. IEEE Transactions in Medical Imaging, 19:1053–1063.

BOOKSTEIN, F. L. 1984. A statistical method for biological shape comparisons. Journal of Theoretical Biology, 107:475-520.

BOOKSTEIN, F. L. 1986. Size and shape spaces for landmark data in two dimensions. Statistical Science, 1:181-242.

BOOKSTEIN, F. L. 1989. Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11:567-585.

BOOKSTEIN, F. L. 1991. Morphometric Tools for Landmark Data: Geometry and Biology. Cambridge Univeristy Press, 435 p.

BOOKSTEIN, F. L. 1996. Standard formula for the uniform shape component in landmark data, p. 153-168. In L. F. Marcus, M. Corti, A. Loy, G. J. P. Naylor, and D. E. Slice (eds.), Advances in Morphometrics. NATO ASI Series A: Life Sciences Volume 284. Plenum Press, New York.

BOOKSTEIN, F. L. 1997. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. Medical Image Analysis, 1:97-118.

BOOKSTEIN, F. L., A. P. STREISSGUTH, P. D. SAMPSON, P. D. CONNOR, AND H. M. BARR. 2002. Corpus callosum shape and neuropsychological deficits in adult males with heavy fetal alcohol exposure. Neuroimage, 15:233-251.

CHAPMAN, R. E. 1990. Conventional Procrustes approaches, p. 251-267. In F. J. Rohlf, and F. L. Bookstein (eds.), Proceedings of the Michigan Morphometrics Workshop. University of Michigan Museum of Zoology Special Publication Number 2, Ann Arbor, Michigan, 380 p.

CHATFIELD, C., AND A. J. COLLINS. 1980. Introduction to Multivariate Analysis. Chapman and Hall/CRC, Boca Raton, Florida, 246 p.

DRYDEN, I. L., AND K. V. MARDIA. 1998. Statistical Shape Analysis. John Wiley and Sons, Chichester, England, 347 p.

FOOTE, M. 1993. Contributions of individual taxa to overall morphological disparity. Paleobiology, 19:403-419.

GOODALL, C. 1991. Procrustes methods in the statistical analysis of shape. Journal of the Royal Statistical Society, Series B (Methodological), 53:285-339.

GOWER, J. C. 1975. Generalized Procrustes analysis. Psychometrika, 40:33-50.

GREEN, W. D. K. 1996. The thin-plate spline and images with curving features, p. 79-87. In K. V. Mardia, C. A. Gill and I. L. Dryden (eds.), Image Fusion and Shape Variability. University of Leeds Press, Leeds.

GUNZ, P., P. MITTEROECKER, AND F. L. BOOKSTEIN. 2005. Semilandmarks in three dimensions, p. 73–98. In D. E. Slice (ed.), Modern Morphometrics In Physical Anthropology. Kluwer Academic Publishers/Plenum, New York.

KENDALL, D. G. 1977. The diffusion of shape. Advances in Applied Probability, 9:428-430.

KENDALL, D. G. 1984. Shape manifolds, Procrustean metrics, and complex projective spaces. Bulletin of the London Mathematical Society, 16:81-121.

KIM, K., H. D. SHEETS, R. A. HANEY, AND C. E. MITCHELL. 2002. Morphometric analysis of ontogeny and allometry of the Middle Ordovician trilobite *Triarthrus becki*. Paleobiology, 28:364-377.

LELE, S. 1991. Some comments on coordinate-free and scale-invariant methods in morphometrics. American Journal of Physical Anthropology, 85:407-417.

LELE, S. 1993. Euclidean Distance Matrix Analysis (EDMA): Estimation of Mean Form and Mean Form Difference. Mathematical Geology, 25:573-602.

LELE, S., AND J. T. RICHTSMEIER. 1991. Euclidean distance matrix analysis: A coordinate-free approach for comparing biological shapes using landmark data. American Journal of Physical Anthropology, 86:415-427.

MARDIA, K. V., J. T. KENT, AND J. M. BIBBY. 1979. Multivariate Analysis. Academic Press, London.

O'HIGGINS, P. 2000. The study of morphological variation in the hominid fossil record: biology, landmarks and geometry. Journal of Anatomy, 197:103-120.

PRESS, W. H., B. P. FLANNERY, S. A. TEUKOLSKY, AND W. T. VETTERLING. 1988. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press.

RICHTSMEIER, J. T., AND S. LELE. 1993. A coordinate-free approach to the analysis of growth patterns: models and theoretical considerations. Biological Reviews, 68:381-411.

RICHTSMEIER, J. T., V. B. DELEON, AND S. R. LELE. 2002. The promise of geometric morphometrics. Yearbook of Physical Anthropology, 45:63-91.

ROHLF, F. J. 1990. Rotational fit (Procrustes) methods, p. 227-236. In F. J. Rohlf, and F. L. Bookstein (eds.), Proceedings of the Michigan Morphometrics Workshop. University of Michigan Museum of Zoology Special Publication Number 2, Ann Arbor, Michigan, 380 p.

ROHLF, F. J. 1998. On applications of geometric morphometrics to studies of ontogeny and phylogeny. Systematic Biology, 47:147-158.

ROHLF, F. J. 1999. Shape statistics: Procrustes superimpositions and tangent spaces. Journal of Classification, 16:197-223.

ROHLF, F. J. 2000. On the use of shape spaces to compare morphometric methods. Hystrix, 11:9-25.

ROHLF, F. J. 2003a. tpsSmall. Version 1.20. Department of Ecology and Evolution, State University of New York. Available at <http://life.bio.sunysb.edu.morph/>.

ROHLF, F. J. 2003b. Bias and error in estimates of mean shape in geometric morphometrics. Journal of Human Evolution, 44:665-683.

ROHLF, F. J. 2008. tpsRelW. Version 1.46. Department of Ecology and Evolution, State University of New York. Available at <http://life.bio.sunysb.edu.morph/>.

ROHLF, F. J. 2009. tpsDig. Version 2.14. Department of Ecology and Evolution, State University of New York. Available at <http://life.bio.sunysb.edu.morph/>.

ROHLF, F. J., AND D. SLICE. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Systematic Zoology, 39:40-59.

SAMPSON, P. D., F. L. BOOKSTEIN, H. SHEEHAN, AND E. L. BOLSON. 1996. Eigenshape analysis of left ventricular outlines from contrast ventriculograms, p. 131-152. In L. F. Marcus, M. Corti, A. Loy, G. J. P. Naylor and D. E. Slice (eds.), Advances in Morphometrics. Nato ASI Series, Series A: Life Science, New York.

SHEETS, H. D., K. KIM, AND C. E. MITCHELL. 2004. A combined landmark and outline-based approach to ontogenetic shape change in the Ordovician trilobite *Triarthrus becki*, p. 67-82. In Elewa, A. M. T. (ed.), Morphometrics: Applications in Biology and Paleontology. Springer-Verlag, Berlin.

SHEETS, H. D., K. M. COVINO, J. M. PANASIEWICZ, AND S. R. MORRIS. 2006. Comparison of geometric morphometric outline methods in the discrimination of age-related differences in feather shape. Frontiers in Zoology, http://www.frontiersinzoology.com/content/3/1/15.

SIEGEL, A. F., AND R. H. BENSON. 1982. A robust comparison of biological shapes. Biometrics, 38:341-350.

SLICE, D. E. 2007. Geometric morphometrics. Annual Review of Anthropology, 36:261-281.

SMALL, C. G. 1996. The Statistical Theory of Shape. Springer Series in Statistics. Springer, New York, 227 p.

WEBSTER, M., H. D. SHEETS, AND N. C. HUGHES. 2001. Allometric patterning in trilobite ontogeny: testing for heterochrony in *Nephrolenellus*, p. 105-144. In M. L. Zelditch (ed.), Beyond Heterochrony: The Evolution of Development. Wiley and Sons, New York.

ZELDITCH, M. L., D. L. SWIDERSKI, H. D. SHEETS, AND W. L. FINK. 2004. Geometric Morphometrics of Biologists: A Primer. Elsevier Academic Press, San Diego, 443 p.