RESEARCH ARTICLE

Inferring the Nature of Allometry from Geometric Data

Kim van der Linde · David Houle

Abstract The form of an organism is the combination of its size and its shape. For a sample of forms, biologists wish to characterize both mean form and the variation in form. For geometric data, where form is characterized as the spatial locations of homologous points, the first step in analysis superimposes the forms, which requires an assumption about what measure of size is appropriate. Geometric morphometrics adopts centroid size as the natural measure of size, and assumes that variation around the mean form is isometric with size. These assumptions limit the interpretation of the resulting estimates of mean and variance in form. We illustrate these problems using allometric variation in shape. We show that superimposition based on subsets of relatively isometric points can yield superior inferences about the overall pattern of variation. We propose and demonstrate two superimposition techniques based on this idea. In subset superimposition, landmarks are progressively discarded from the data used for superimposition if they result in significant decreases in the variation among the remaining landmarks. In outline superimposition, regularly distributed pseudolandmarks on the continuous outline of a form are used as the basis for superimposition of the landmarks contained within it. Simulations show that these techniques can result in dramatic improvements in the accuracy of estimated variance-covariance matrices among landmarks when our assumptions are roughly satisfied. The pattern of variation inferred by means of our superimposition techniques can be quite different from that recovered from full generalized Procrustes superimposition. The pattern of shape variation in the wings of drosophilid flies appears to meet these assumptions. Adoption of superimposition procedures that incorporate biological assumptions about the nature of size and of the variation in shape can dramatically improve the ability to infer the pattern of variation in geometric morphometric data.

The form of an organism is its combined size and shape. By definition, shape consists of those aspects of form that remain when size is removed (e.g., Mosimann 1970). A wide variety of biological problems concern the evolution of mean form and the role of variation in form in evolution, genetics, development and disease. Biologists are interested in studying the causes and consequences of variation in form. For example, we would like to detect whether shape changes with size (that is whether shape is allometric), and more interestingly describe what it is about shape that changes with size. Alternatively we might want to study changes in shape due to a genetic difference. The first question again is whether form differs between genotypes, but once that is answered it would be of great value to be able to infer which parts of form are affected by a genetic change.

In this paper we call attention to the fact that while the first question—Is there variation?—can be answered unambiguously with current approaches, the second—What about shape varies?—cannot. Traditional multivariate analyses do not take advantage of the spatial information in

K. van der Linde (✉) · D. Houle
Department of Biological Science, Florida State University, 319 Stadium Drive, Tallahassee, FL 32306-4295, USA
e-mail: kim@kimvdlinde.com

D. Houle
e-mail: dhoule@bio.fsu.edu

the data, or are confounded by it. Methods that do take the spatial nature of the variation into account, geometric morphometrics (Bookstein 1991, 1996; Rohlf and Marcus 1993; Adams et al. 2004; Zelditch et al. 2004) or Euclidean Distance Matrix Analysis; EDMA (Lele 1993; Lele and Richtsmeier 2001; Lele and McCulloch 2002; Richtsmeier et al. 2005), can answer the first question but are technically blind to the second.

The fundamental source of this limitation is that the spatial data at the heart of morphometrics do not come with a natural coordinate system common to all specimens. The data themselves must be used in some fashion to infer a common coordinate system, and this process certainly confounds variation at different homologous points in the form (Bookstein 1986; Rohlf and Slice 1990; Goodall 1991; Lele 1993; Lele and McCulloch 2002; Richtsmeier et al. 2005). The result is that morphometric data is only considered to be suited to testing hypotheses about mean form—for example, 'Do two groups differ in shape?', or 'Does mean shape change in a consistent way with a covariate? Hypotheses about the nature of variation in form, such as the relative variability of different parts of the form, are considered to be outside the purview of statistical tests in geometric morphometrics because estimating rotation and translation, the 'nuisance' parameters, plus scaling or size requires four degrees of freedom for planar data, but information about every point is still required for estimation of the covariance matrix, which is therefore not fully estimable without a priori information. This is known as the identifiability problem (Lele 1993; Lele and McCulloch 2002).

Despite these widely known difficulties, the intuition is widespread that the patterns of covariance revealed by geometric morphometrics really are meaningful, at least as a tool for hypothesis generation. A compelling sign of this is the widespread utilization of visualizations of variations in particular directions in morphospace. If that intuition is taken seriously analyses should be developed that take advantage of the real information that is present, even if it is confounded to some extent by the superimposition algorithm used.

Furthermore, just as we know that geometric morphometric techniques cannot in general recover the true pattern of variation and covariation, we know that they can do so for the particular pattern of variation that corresponds to the assumptions that are employed in the analysis. The most widely used superimposition approach is generalized Procrustes superimposition (GPS), in which the set of nuisance parameters is chosen to minimize the sum of squared distances between the points of each configuration and a mean form, subject to a particular definition of scaling (Gower 1975; Rohlf and Slice 1990). While this choice of a least-squares criterion suggests that the goal of

this transformation is statistical in nature, it is justified instead because it places the landmark configurations in a particular space of all possible rearrangements of k points in d dimensions. This space is related to Kendall shape space (Slice 2001), and has a convenient geometry for some kinds of questions. Nevertheless, the pattern of variation among landmarks following GPS will accurately reflect the pattern of variation if the variation in form reflected in each point is isotropic (equal and uncorrelated) (Dryden and Mardia 1998).

This suggests that if appropriate a priori assumptions about the pattern of variation are made the identifiability problem can be partly or wholly overcome (Goodall 1995), and the pattern of variation interpreted. Some attempts have been made to move in that direction. Goodall (1991, 1995) proposed estimation of variance-covariance matrices by iterative superimposition based on Mahalanobis (variance weighted) distance, rather than Euclidean distance. More recently, a maximum likelihood approach has been implemented for 3-D data using constraints on the pattern of covariation around each landmark, and the distribution of size and the eigenvalues (Theobald and Wuttke 2006).

We will not solve this problem here. Our purpose is first to call attention to the problem. Second we propose two related exploratory approaches that modify the GPS approach based on simple biological assumptions about the nature of variation in form. These can improve our ability to draw inferences about changes in form if our assumptions about the way that form changes are correct.

The Problem

A Simple Example

We start by examining a simple data set that contains one landmark that varies allometrically as well as four that are isometric. Consider a set of square forms of different sizes, where the four corners are used as landmarks. In addition, we place one landmark along the outline of the square at a constant distance from one of its corners, as shown in Fig. 1a, ensuring that the relative position of this fifth point changes relative to the other points; that is, it is allometric. Because the true center of the form is equidistant from the four corners, a 'biological' measure of size is based on just the four corners. With this true model for size and variation shape in mind, we can superimpose the forms using generalized Procrustes superimposition (GPS) on just the corner landmarks. The scaling, translation and rotation parameters fit from these four points are then used to place the fifth point. This superimposition recovers a pattern of variation that reflects the model we used to generate the data, as shown in Fig. 1c; the four corner landmarks align
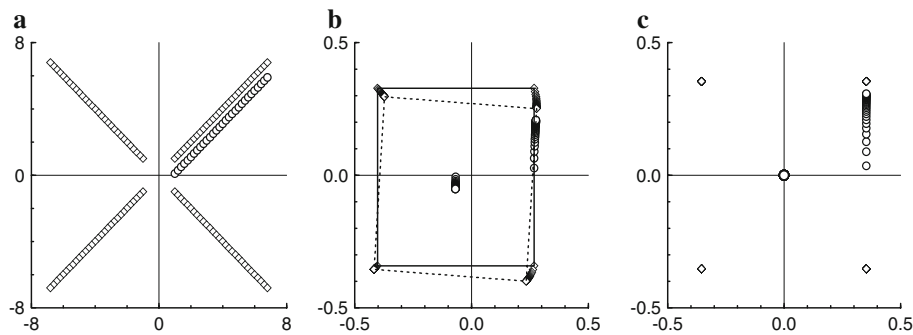
Fig. 1 Five-landmark test data set. **a** Data as generated. The variation in the corner points (*squares*) is isometric, whereas the variation in the last point (*circles*) is allometric. Centroid size ranges from 1 to 6.8. **b** Generated data superimposed by GPS on the basis of all five points.

The solid line is the largest individual, the dashed line the smallest. The circles in the middle represent the positions of the centers of the squares. **c** Generated data superimposed on the basis of the four isometric corner points

perfectly, and all of the shape variation is in the fifth point. Because the true 'natural' center of the form is equidistant from the four corners, centroid size calculated based on just the four corner points corresponds to an intuitive measure of linear size.

The results of GPS of all five points suggest quite a different pattern of variation. Figure 1b shows the GPS superimposed data, with the outlines of the largest and smallest square and the midpoints of all squares drawn in. Most noteworthy is that all five landmarks now show substantial variation due to the spreading out of variation present in the fifth landmark. The addition of the fifth landmark has affected the estimation of all three sets of nuisance parameters. The effect on centering and rotation are immediately clear from the figure. Less obvious is that the sizes of the forms are also affected. When centroid sizes are adjusted to the same value, the square with the largest area is shown as smaller, whereas the originally smallest square is now the largest, because the pattern of allometry locates the fifth landmark relatively farther from the centroid as size increases.

The behavior of these data under GPS is a function of several well-known properties. The first is the inability of general superimposition algorithms to recover a known covariance structure because of the identifiability problem. Second, the intuitive center of the object (the center of the square) is not located at the centroid, the average position of the landmarks. Third the intuitive definition of size based on the overall size of the object (in this case a dimensions of the square) differs from centroid size. The last two issues are one important justification for the advice to find a set of landmarks that are well spaced on the form (Bookstein 1991).

Note that even if we did not know the true model for the relationship between shape and size (that is, we did not have access to Fig. 1a), two aspects of the data could point us in the direction of the true, underlying model. First, even

with the five-point-superimposed data, the fifth point is clearly more variable than the others. Second, if the area or outlines of the forms have some biological meaning, superimposition based on just the outlines will be intuitively appealing. Goodall (1991) perceptively noted that "data come with a perturbation to the coordinate frame, but also with clues, at least partially independent of the landmark data, to the correct registration. An example of such clues is the outline of bones, between the landmarks which are at the sutures of the bones."

Allometry and Superimposition in Drosophilid Wings

We next consider a set of real data on the form of drosophilid wings that shows patterns of variation that we believe are similar to those in the simple example above. A typical *Drosophila* wing is shown in Fig. 2a. The topology of five longitudinal veins (the first being the leading edge of the wing) and two crossveins is almost invariant in the family Drosophilidae. We use a semi-automated system to fit a series of B-splines to the vein structure (Houle et al. 2003), as shown in Fig. 2b. Our previous work in this system (e.g., Mezey and Houle 2005) has been based on the set of 12 vein or edge/vein landmarks shown in Fig. 2a.

We have now ported the WINGMACHINE software described in Houle et al. (2003) to Java (Sun Microsystems Inc. 1992–2006) and enhanced it to allow us to edit the splines directly; the landmarks are extracted from the spline data (van der Linde and Houle 2004–2008). In implementing the techniques described below, we have developed Geometrics (van der Linde 2005–2008) and Spline packages in Java, based on published algorithms (Siegel and Benson 1982; Rohlf and Slice 1990; Goodall 1991; Akca 2003; Rohlf and Bookstein 2003).

To investigate the superimposition issues that arise as a result of allometric variation in drosophilid wings, we haphazardly chose, from a much more extensive data set,
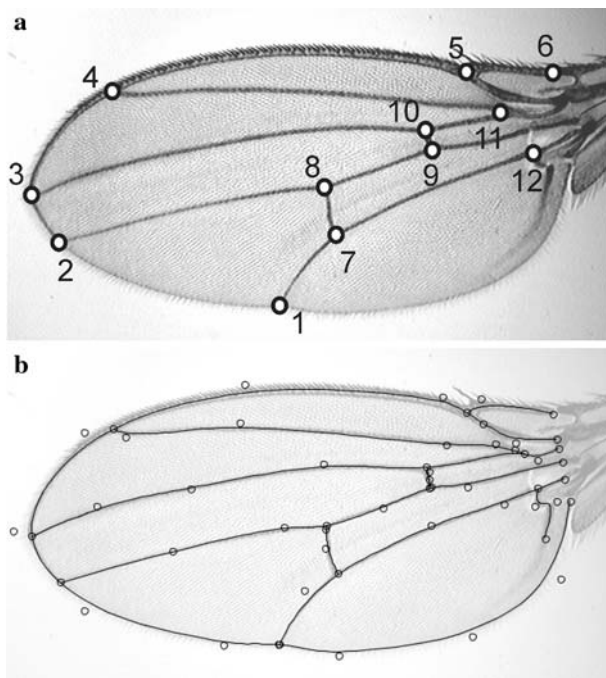
Fig. 2 A typical *Drosophila* wing. Analysed landmarks (*top*) and clamped quadratic B-splines with control points (*bottom*)

four species that differ widely in mean centroid size (CS): *Scaptodrosophila dorsocentralis* (mean CS = 1.44 mm), *Drosophila (Sophophora) equinoxialis* (CS = 1.96 mm), *Zaprionus ghesquierei* (CS = 2.73 mm), and *Drosophila (Idiomyia) soonae* (CS = 3.98 mm). *Drosophila soonae* was obtained from the *Drosophila* species stock center; *D. equinoxialis* was collected in Panama by D.H.; *S. dorsocentralis* and *Z. ghesquierei* stocks were furnished by Jean David. All data were obtained from lab-reared flies. Twenty-five individuals of each species were used in the following analyses.

The GPS superimposed data are shown in Fig. 3a. Several aspects of this superimposed configuration are noteworthy. First, the amount of variation inferred for each landmark varies markedly. In particular, landmark 4 at the upper left of the figure has far more variance than the other landmarks, to the point where the among-species variation is great enough to result in four discrete clusters of points. Some of this pattern is caused by the clumping of landmarks at the proximal end of the wing, on the right side of Fig. 3a, but note that landmarks 2 and 3, which are even more distal than 4, have far less variation. Landmarks 7 and 8 are also noticeably more variable than the other landmarks. Second, the species means for all the landmarks show a directional change correlated with size, implying existence of allometry in the original point configurations. This allometry is confirmed by regression of shape on centroid size (performed in Rohlf 2003), which

convincingly rejects the null hypothesis of isometry (Wilk's $\Lambda = 0.012$, df 20, 68, $P = 0.001$ by permutation test). Third, the areas enclosed within the outlines of the species mean wing shape differ strikingly. As with the simulated data, the shape space inferred by GPS results in the comparison of species at different adjusted wing areas, and one that reverses the size relationships among species. Clearly the pattern of allometry is such that at least some of the central landmarks are more displaced from the centroid in larger species.

Dealing with the Problem

Departures from Isometry

Current morphometric practice deals with the situation shown in Fig. 3a in several ways. First, one might be content with the verbal description of what one can infer about how shape changes with size, given one's knowledge or intuition about the properties of GPS, such as we have constructed above. To this description, one can add resuperimpositions of the data using other algorithms, such as a resistant (Rohlf and Slice 1990; Walker 2000) or two-point registration, with more interpretation. Third, one might instead analyze these data with a superimposition-free approach, such as EDMA (Lele and Richtsmeier 2001; Richtsmeier et al. 2005), which forgoes much of the opportunity to visualize and interpret the pattern of covariation. Our major interest is to determine the causes of variation in shape, at both the developmental and the evolutionary levels. A crucial step in doing so is to generate hypotheses about the true pattern of variation and covariation in form.

In the Introduction we noted that GPS is constructed to place morphometric data in a geometric space with particular properties and is thus not considered by its leading proponents to be an estimation procedure. We and others (Theobald and Wuttke 2006) explicitly break with this tradition because our goal is to estimate the variance-covariance matrix of variation in form—that is estimation. For this purpose we need to make assumptions about some aspects of that variation, which we refer to as biological assumptions. We want to adapt the GPS approach, insofar as possible, to our inference problem. In this new context, GPS uses assumptions that can be interpreted as biological assumptions—namely that variation in isotropic. Here we try to adapt the machinery of GPS to mitigate the impact of this clear departure of GPS from biologically motivated assumptions. Biological variation is certainly not isotropic.

When the assumptions of the superimposition algorithm correspond to the actual distribution of the data, the known pattern of variance and covariance among points can be recovered. Our proposed approach is then to seek the

Fig. 3 Superimpositions of individuals by various techniques. The solid line (*solid circles*) represents the average *Scaptodrosophila dorsocentralis* (the smallest species); the dashed line (*open triangles*) represents *Drosophila soonae,* the largest species. The centroid is at the junction of the horizontal and vertical gridlines. **a** Superimposition by traditional GPS. **b** Superimposition by traditional GPS, but with landmarks 4, 7, and 8 omitted. **c** Superimposition by generalized procrustes analysis with 12 pseudolandmarks representing the outline. **d** Superimposition by generalized oblique analysis (GPS with affine component) with 12 pseudolandmarks representing the outline
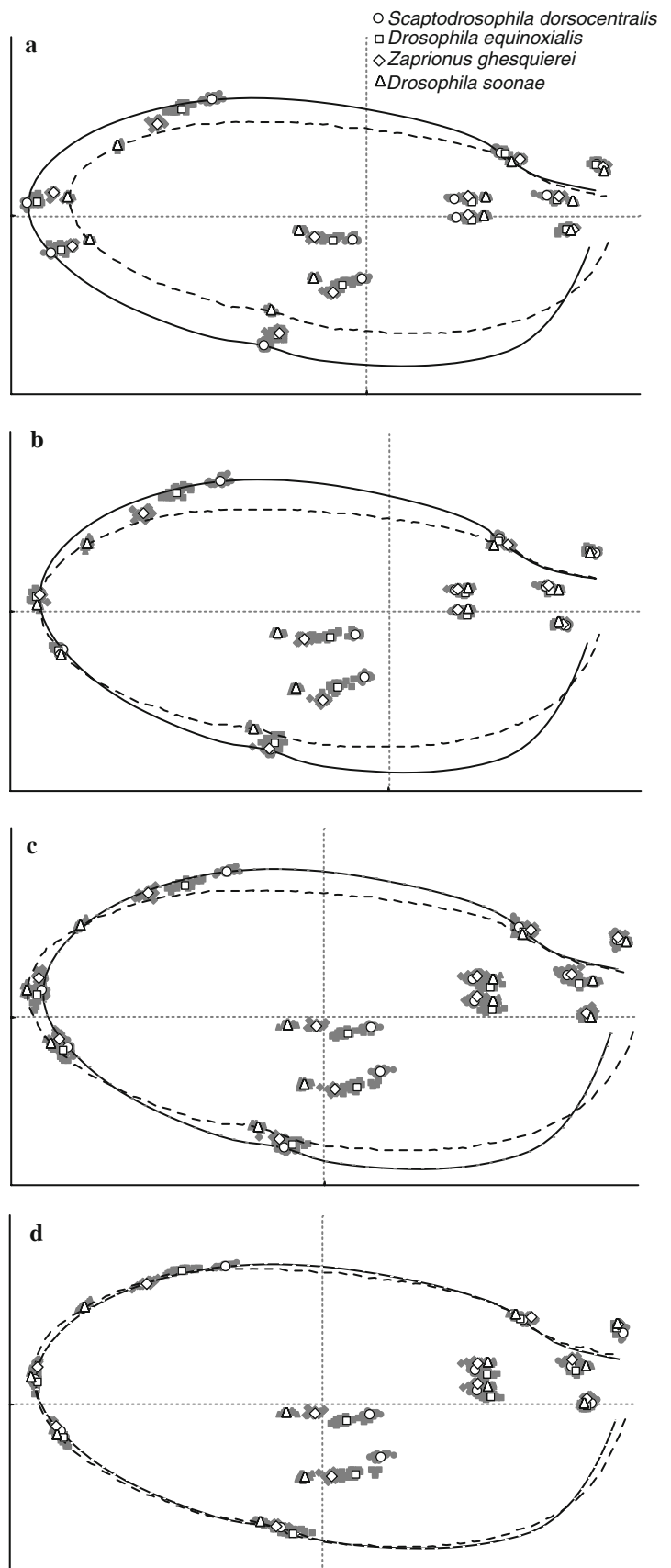
subset of information about the form that corresponds best to the assumptions of GPS, then to superimpose all of the data on the basis of this subset of points. This would be fully justified under the following assumptions. First, we assume that there are a subset of landmarks that have isotropic variation with respect to each other. This might be the case if digitizing error is the only source of variation in that subset. Second, we assume that other landmarks have, in addition to isotropic error, additional biological variation, perhaps due to allometry. Thus those points with the lowest inferred variation are most likely to correspond to those with isotropic error, and could correspond to the assumptions of GPS. When allometry is demonstrated, as in the fly wing data, we exclude from the superimposition those points that show the highest level of variation in shape space. The result is analogous to the use of two-point registration (Bookstein coordinates), but where the number of points used may be more than two.

We hasten to point out that for biological data the true pattern of covariance is unknown and that no subset of points need correspond to the assumptions of GPS. Results recovered using our approach are exploratory. A key issue is what this method recovers when the full set of assumptions above in not justified, for example, if all points are allometric, but some are more allometric than others. We predict that our procedure should lead to improved inferences about the source of variation in shape in such cases, an assertion we investigate with simulated data below.

We discuss two variants of this basic approach. The first is a simple algorithm for determining which of the landmark points should be excluded from the set of points used. We call it subset superimposition. Second, we propose that, when data about the outline of the form are also available, using the outline as the basis for superimposition may substantially improve our estimate of the covariances among the landmarks. We call this procedure outline superimposition.

Subset Superimposition

Under a null hypothesis of isometry, where all landmarks have the same variance-covariance pattern and the same relationship with size, each landmark is expected to contribute equally to the residual Procrustes sum of squares (PSS). Conversely, if a particular landmark is more allometric than the remainder, then superimposition of the remainder without reference to the allometric point will result in an improved fit of the overall superimposition. This process can be repeated until the user judges that a more satisfactory superimposition is obtained. It is easy to imagine how this process proceeds for the simulated data in Fig. 1. In the absence of knowledge about the true model,

one would superimpose these data using GPS of all five points, producing the inferred pattern in Fig. 1b. The results clearly show that the non-corner landmark has far more variance than any of the other points. This point is then removed from set used to estimate the nuisance parameters, and GPS repeated. This results in Fig. 1c, in this case an obviously superior superimposition. In cases where a subset of the landmarks follows a common pattern of variation, and where this pattern seems to reflect the size of the form accurately, the following algorithm will yield an improved superimposition.

Algorithm: Subset Superimposition

Consider data on n configurations in the positions of k landmarks in 2 dimensions. Arrange these coordinates into n $k \times 2$ configuration matrices $\mathbf{X}_{i-0}$. Landmarks are removed sequentially, so at the mth iteration, only $k - m$ points remain, and the configuration matrix $\mathbf{X}_{i-m}$ used for superimposition has dimension $k - m \times 2$. Each iteration consists of the following steps:

1. Perform a generalized Procrustes superimposition of the forms $\mathbf{X}_{i-m}$, to obtain the superimposed, scaled, configurations $\mathbf{X}_{i-m}^*$ and a mean or tangent shape $\bar{\mathbf{X}}_m^*$.
2. Decompose the Procrustes sum of squares (PSS) into the parts due to each of the $k - m$ landmarks. The sum of squares for the jth landmark is the jth diagonal element of the point-wise SSCP matrix $\mathbf{D}^2 = \sum_{i=1}^{n} \left[\left(\mathbf{X_i^*} - \bar{\mathbf{X}}^* \right) \left(\mathbf{X_i^*} - \bar{\mathbf{X}}^* \right)^t \right]$, termed D_j^2. The trace of $\mathbf{D}^2$ is the overall PSS.
3. Examine the distribution of the PSS values to determine whether removing a point is justified. We discuss possible criteria below.
4. Identify the most deviant D_j^2, then remove the jth row from all of the $\mathbf{X}_{i-m}$ to yield configurations $\mathbf{X}_{i-(m+1)}$. Increment m and return to step 1.
5. Once the stopping rule in step 3 is satisfied, the original configuration matrices $\mathbf{X}_{i-0}$ are scaled, translated, and rotated with the nuisance parameters estimated during the final Procrustes superimposition step on the configurations $\mathbf{X}_{i-m}$ to yield the final superimposed and scaled configurations $\mathbf{X}_i^*$ and tangent shape $\bar{\mathbf{X}}^*$.

There are several alternative choices for stopping criteria for step 3. The first alternative is an intuitive one, based on changes in PSS as a function of deletion step, m. This yields a plot analogous to a scree plot of eigenvalues, as shown in Fig. 4 (dashed line). With our selection algorithm, PSS must decrease with each successive deletion of a point. This plot gives an indication of the change in variance, potentially representing a gain in isometry, with each successive deletion. For the drosophilid data in

Fig. 4 Residual Procrustes sum of squares for the *Drosophila* data as points are sequentially deleted from the superimposition. Solid line, log-transformed values (*left* y-axis); dashed line, untransformed values (*right* y-axis)
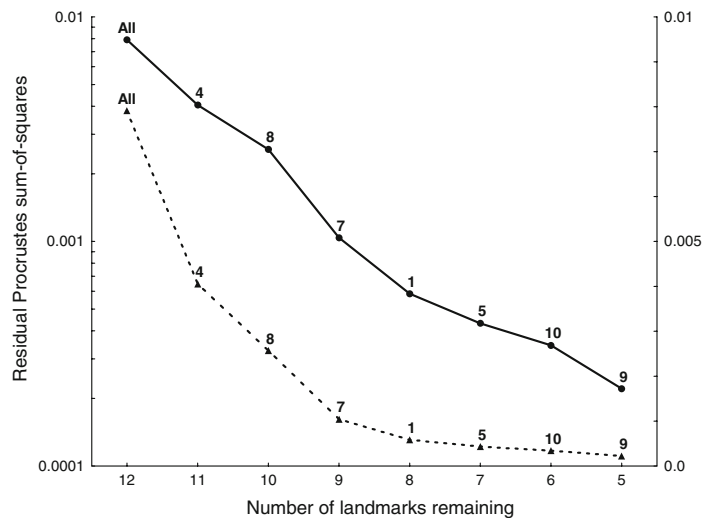

Fig. 4, deletion of points 4, 7, and 8 results in a very large reduction in PSS. The resulting superimposition with points 4, 7, and 8 omitted is shown in Fig. 3b.

A potential disadvantage of this intuitive approach is that users may feel free to choose the degree of deletion to yield particular results. To avoid this, it might be better to choose an algorithmic approach *a priori*. One algorithmic approach could be based on statistical testing of the underlying homogeneity of variances assumption of Procrustes superimposition. Under the null hypothesis, $D_1^2 = D_2^2 = \ldots = D_k^2$. We propose using the results of a Levene test for equality of these variances. When the test is significant, one deletes an additional point in step 4 of the algorithm. If the test is not significant, one obtains the final estimates of shape in step 5 of the algorithm. This approach may ultimately lead to a large number of statistical tests, which can be partially compensated for by dividing the critical P value of each test by the number of tests $(m + 1)$. This alone will not yield correct probabilities, as subsequent tests are not independent. The value of having an algorithmic stopping rule may outweigh the statistical uncertainty. A potential problem with this stopping rule is that the null hypothesis of equal Procrustes sums of squares (PSS) at each point is not necessarily met for any pair of points. An alternative stopping rule is to complete the deletion algorithm down to some arbitrary fraction of points, such as one half, with no statistical testing.

We applied this algorithm to the *Drosophila* data, and Fig. 4 shows PSS among the landmarks used at each iteration of the analysis. Levene tests remained significant at each step, so that stopping criterion was never satisfied. The reason is clear on a log scale, where the decrease in PSS is approximately linear (Fig. 4, *solid line*).

Performance of Subset Alignment for Simulated Data

We investigated the properties of subset removal on simulated data for which the correct model is known. The data were modeled on the basis of the change in a reference form **R** (of dimension $k \times 2$) as a function of an underlying size variable S. The mth dimension of landmark j in individual i is

$$X_{ijm} = R_{jm}S_i^{A_{jm}} + B_{jm} + E_{ijm}S_i$$

where A is a matrix of allometric coefficients, B is a matrix of constant deviations from the reference (as in point 5 in Fig. 1), and Ei is a matrix of multivariate normal deviations with distribution $N(0, \Sigma)$. This procedure generates ideal data where the nuisance parameters of translation and rotation are absent.

We then calculated estimates of the variance covariance matrix of the landmarks using different superimpositions. We estimate the true model covariance matrix, P, after scaling the configurations by the parametric size $S_i^{-1}\mathbf{X_i}$, which is equivalent to using a superimposition based only on isometric points. To assess the performance of superimposition algorithms, we subjected the unscaled forms either to simple GPS or to the subset superimposition algorithm, then estimated the covariance matrix $\hat{\mathbf{P}}$. To judge the performance of the superimposition algorithm, we estimated the scaled mean square error of $\hat{\mathbf{P}}$ as

$$MSE = \frac{\sqrt{\sum_{i=1}^{k} \sum_{j=i}^{k} \left[\left(p_{ij} - \hat{p}_{ij} \right)^2 \right]}}{\bar{p}_{ij}\, k(k+1)/2},$$

where $\bar{p}_{ij}$ is the mean of the unique elements of P. Under a perfect superimposition, MSE reaches zero.

In our simulations we allowed subsets of $n < k$ points to show different degrees of allometry, while the remaining
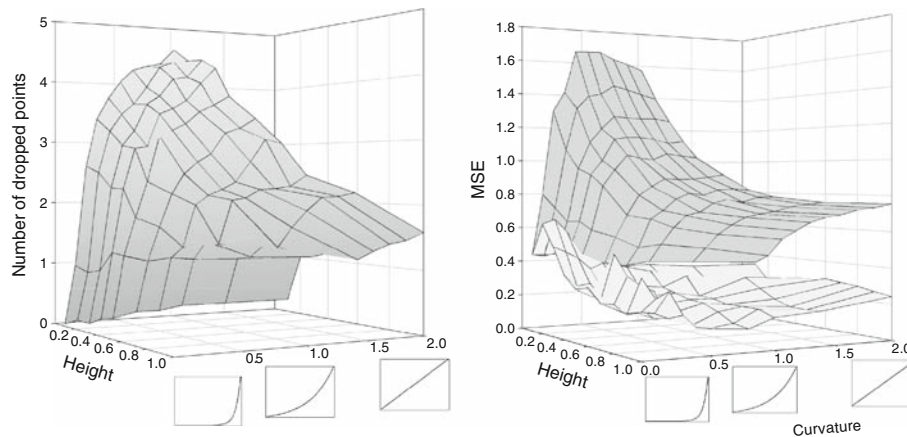
Fig. 5 Results of the simulation using four isometric and four positively allometric points. Height at the x-axis indicates the magnitude of the allometry; curvature at the y-axis indicates the degree to which the allometry increases between points (0 is linear, 2 is most curved). **a** z-axis indicates how many points are dropped from the superimposition. **b** z-axis values are the mean square error for all points. The dark-gray plane indicates the original MSE; the light-gray plane indicates the residual MSE after deletion of the points that are determined to be deviating from isotropy

$k - n$ points were isometric. The increment in the allometry of each point varied from a and incremented the degree of allometry. In each set of simulations, we constructed a $k \times 2$ matrix A of allometric coefficients a_{ij}. For simplicity, we assumed that the x and y coefficients were equal ($a_{i1} = a_{i2}$). The degree of allometry of the ith point is calculated as

$$a_{ij} = t + h \left(\frac{e^{i * curvature} - 1}{e^{n * curvature} - 1} \right) \quad \text{for } i \leq n \text{ and}$$

$$a_{ij} = t \quad \text{for } i > n$$

where h indicates the maximum difference in allometry between the points. When $t = 1$, points with $i > n$ are isometric. The curvature parameter controls the rate at which allometry changes with i. When curvature is near 0, the increment is linear in i. As curvature increases away from 0, the allometry is an accelerating function of i.

Our first set of simulations was of situations where four of the eight points were isometric ($t = 1$, $k = 8$, and $n = 4$), and subset superimposition was expected to perform well. In each simulation, we chose 30 different sizes uniformly distributed between 1 and 3.9. For each parameter combination, we simulated 50 configurations. Figure 5a shows the number of points that are dropped when the Levene test criterion is used for various values of height and curvature, average of 50 replicas per combination of height and curvature. When the points are clearly allometric, the algorithm drops all of the allometric points, but this outcome becomes less likely as the degree of allometry becomes smaller. Figure 5b compares the MSE of the inferred variance covariance matrices for the full set of points under GPS and after application of the subset algorithm with the Levene test as a stopping rule. Subsetting never results in a worse fit to the true covariance matrix,

and it leads to a dramatic increase in accuracy when the variance in allometry of the points is large.

In our second set of simulations, we modeled a worst-case scenario for our algorithm, where all points depart from isometry, half with allometric coefficients greater than one and half less than one, corresponding to a set of landmarks that do not represent overall size well. In these cases, MSE generally increased after subsetting (Fig. 6).

To summarize, subset alignment uses the relative contribution to the Procrustes sum of square of each landmark to determine which landmark to drop from the analysis, in order to obtain a subset of landmarks that potentially approximates the isotopic error distribution assumed by the Procrustes analysis. The simulations showed that this algorithm improved estimates of the variance-covariance matrix when the relative contribution to the Procrustes sum of squares differed between points, the isometric points contributed less variation. As expected, when the relative contribution to the Procrustes sum of squares is roughly equal for each landmark, no such clear improvement occurred. Note that our algorithm could retain a subset of covarying allometric points, and through that, result in the dropping of the isometric points when the overall set of landmarks do not adequately reflect size.

Outline Superimposition

Our second proposal for dealing with allometry is to use the outline of the form as the basis for superimposition of the landmarks. This approach was suggested by our fly wing data, where we automatically estimate both outlines and landmarks in the same step, as shown in Fig. 2b. Examination of the wings of the four species in our data set

suggested to us that species varied less in the overall shape of the wing than in the arrangement of the landmarks. To determine whether this was the case, we generated 12 evenly spaced pseudolandmarks along the outline of the longest curve that mostly traces the wing outline. We then used GPS to superimpose the configurations of 12 landmark plus 12 pseudolandmark points simultaneously using either the 12 landmarks (Fig. 3a) or the 12 pseudolandmarks (Fig. 3c) to estimate the nuisance parameters. The squared Procrustes sum of squares of the 12 pseudolandmarks in this shape space was indeed only one-eighth that of the landmarks (9.82×10^{-4} vs 7.86×10^{-3}). This lower Procrustes sum of squares for the outline is not due to geometric constraints. To fit this curve adequately, we use 14 spline control points, so the 12 pseudolandmarks cannot express all of the information in the spline model of this curve. Unlike the pseudolandmarks discussed by Bookstein (e.g., 1991), neither end of this curve is defined by landmark points, so the autocorrelation between landmarks and pseudolandmarks is reduced.

When outline data have this reduced variance relative to landmarks, or where the outline naturally captures biological size, superimposition using only the subset of pseudolandmarks on the outline may be informative. The full set of superimposed configurations can be analyzed as with any subset-superimposed data. An alternative is to discard the pseudolandmarks after superimposition and to analyze only the landmark data. In this case the data used to superimpose the forms are not directly confounded with the data to be analyzed, although of course covariance

between the outline and landmarks will cause dependence between the two. This approach has the additional advantage that the landmark covariance matrix is potentially of full rank.

The results of outline superimposition for the drosophilid data are shown in Fig. 3c. They are similar to those of subset superimposition, in that the variation in the allometric landmarks identified has increased, whereas the variation in the isometric landmarks has been reduced. One attractive property of this superimposition is that the sizes are scaled to match our intuitive notion of wing size based on wing area. The centroid is now close to the biological midpoint of the wing. This result is functionally appropriate, as the lift-generating properties of wings are determined almost entirely by their planar form (Dickinson et al. 1999). Inspection of the mean outlines of each species suggests an allometric relationship between the affine components of the outlines, where smaller species have rounder wings. A model that fits affine parameters can also be incorporated into the superimposition (Rohlf and Slice 1990; Rohlf and Bookstein 2003); the results are shown in Fig. 3d. Incorporation of the affine component into the outline superimposition reduces the PSS by 79% compared to the outline superimposition and by 97% relative to the landmark superimposition.

Simulation results (not shown) confirm that outline superimposition performs much better at recovering the true variance-covariance pattern than does GPS when the outline is isometric and some or all of the landmarks are allometric.

Fig. 6 Effect of simulation using eight allometric points, with both negative allometry and allometry. For details, see Fig. 5
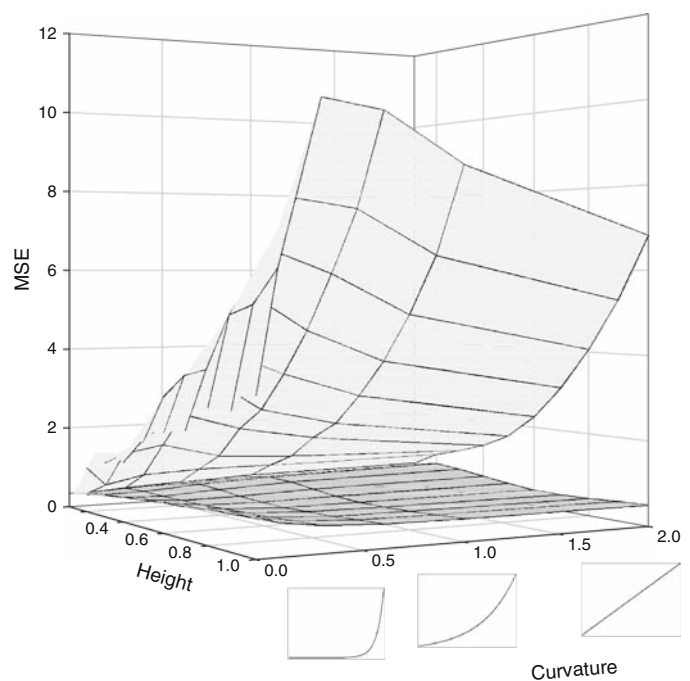
Projections of Kendall's Shape Space

Shapes exist in a curved space called Kendall's shape space, but many statistical methods assume that the data occur in Euclidean space, so determining whether distances between forms in the curved shape space are well approximated by distances in the linearized space tangent to the mean form in the sample is of interest. If these distances are highly correlated, then standard statistical tests are justified. We estimated these correlations for the *Drosophila* data set after GPS, subset, and outline superimpositions. For each superimposition, the R^2 values were essentially perfect ($R^2 = 0.999$ or higher). A similar picture arose from the simulated data, for which we calculated correlations for various extreme cases. Each of these cases resulted in very high correlations whether traditional (0.9926 and higher) or subset superimpositions (0.9939 and higher) were used. In general, superimpositions of the simulated data based on the subset method resulted in higher correlations between shape space and tangent space than those based on all landmarks.

Discussion

As currently practiced, the analysis of shape data by geometric morphometrics is well suited to testing hypotheses about mean shapes but not to consideration of hypotheses about the variation in shape. Many biological questions, such as the causes of variation and the evolution of shape, are therefore outside the purview of current analyses. This situation is clearly not necessary: If one has a model of how variation is produced, it can be used as the basis for morphometric analyses, and its adequacy can be tested in various ways.

A key element in this situation is that the goal of the most widely used analysis of morphometric data, the generalized procrustes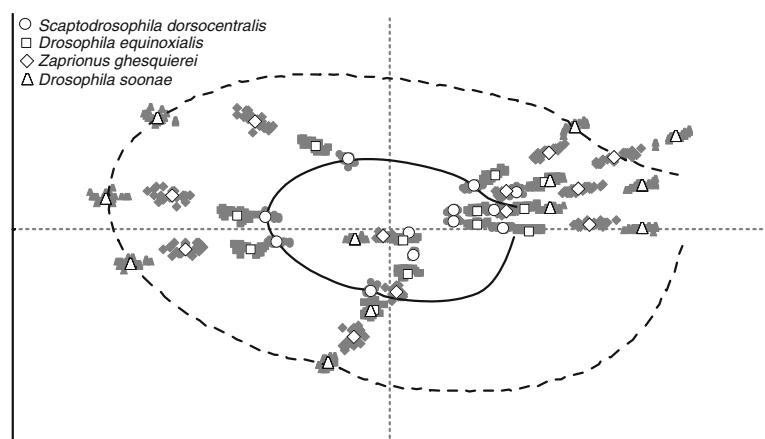 superimposition, is to locate the measured forms in a particular geometric space (Bookstein 1991; Zelditch et al. 2004), rather than to estimate the pattern of variation and covariation in form. To some, this suggests that an entirely new approach needs to be taken if the goal is to estimate variation. While this will likely prove to be true we are interested in the analysis of data now. We have instead proposed modification to the Procrustes approach that can improve its ability to estimate variation under some circumstances.

We have suggested two modifications of Procrustes superimposition techniques that are justified by some simple assumptions about variation in form: that specific parts of form are often strikingly allometric in comparison with the overall form. These relatively allometric parts are detectable by the large residual variation of landmarks in those regions after Procrustes superimposition. When these assumptions are correct, our simulations show that superimposition using a subset of landmarks can dramatically improve our ability to infer the pattern of variation and covariation in shape and therefore help to generate hypotheses about the causes of variation.

Our interest in this problem was sparked by the contrast between the attention paid to detecting allometry in Procrustes-superimposed geometric morphometric analyses and the fact that allometry is a violation of the isotropy assumption implied by the superimposition algorithm. On the other hand, the assumption that the covariation around each landmark is homogeneous used in both EDMA and Theobald and Wuttke's (2006) is clearly not general. In particular, the automated method we use to recover landmark positions on *Drosophila* wings (Houle et al. 2003) makes this assumption unattractive for our data.

Our first proposed technique, subset superimposition, will result in improved inference when landmarks differ in the degree of allometry and the majority of points are isometric. Our second proposed technique, outline superimposition, can result in an improved inference of variation for several reasons. First, it can be considered a special

Fig. 7 Size-restored drosophilid data. Superimposed individuals by generalized oblique analysis (GPS with affine component) with 12 pseudolandmarks representing the outline. Symbols as in Fig. 3

○ *Scaptodrosophila dorsocentralis*
□ *Drosophila equinoxialis*
◇ *Zaprionus ghesquierei*
△ *Drosophila soonae*

case of subset superimposition and is justified if the pattern of variation in the outline is more consistent with isometry than that in the landmarks. Second, in some cases, the outline is the natural locus of the function of the form, so inferring changes in shape relative to the outline of the form will be of biological interest.

The differences between the inferred pattern of covariation obtained with the standard generalized Procrustes (GPS) and that obtained by subset or outline superimposition can be quite large. For example, the implied pattern of variation in the wings of four species of drosophilid flies derived from traditional GPS superimposition, shown in Fig. 3a, differs dramatically from that achieved with subset (Fig. 3b) and outline superimposition (Fig. 3c). Most notably, we can infer that landmarks 1, 4, 7, and 8 show much more allometric variation than is implied in the superimposition based on all the landmarks. Both alternatives result in measures of size that more closely reflect the overall size of the wing blades, a valuable feature given the functional significance of the area of the wing. Figure 7 shows the result after the size component of the shape data was restored. Lines fitted through landmarks 1, 4, 7, and 8 are not going to the centroid, indicating allometry.

These inferred results are interpretable in terms of known aspects of wing development in *Drosophila*. The positions of landmarks 1 and 4 are influenced by variation in the *decapentaplegic* (*dpp*) pathway, which locates veins II and V along the anterior–posterior axis of the wing, relative to an important developmental boundary that runs just posterior to vein III (Held 2002; de Celis 2003). Increased *dpp* activity relative to the rest of the wing would move the intersections of both veins II and V more proximally. The acute angle with which vein II meets the wing margin dictates that landmark 4 will undergo the largest change in position. The determination of the crossveins occurs much later than that of the long veins (Held 2002), so again the positions of landmarks 7 and 8 could plausibly be decoupled from growth of the rest of the wing.

Our proposed modifications of superimposition techniques are not entirely novel. Slice (1998) implemented the possibility of subset superimposition by allowing the user to designate primary points, which are used for superimposition, and secondary points that are not, but we can find no discussion of criteria for choosing such points. Buckley et al. (1999) noted discrepancies in the variances of landmarks after Procrustes superimposition, and discarded those with the largest residuals before analysis.

Resistant-fit superimposition (Siegel and Benson 1982; Rohlf and Slice 1990; Verboon and Gabriel 1995; Dryden and Mardia 1998) can improve superimposition when a distinct minority of points shows anomalous patterns of variation, as in the well-known "Pinocchio case", where one landmark covaries with size very differently from others. Although resistant-fit algorithms do allow detection of such cases (Rohlf and Slice 1990; Walker 2000), in principle such a point-wise difference in variance would be better dealt with by a point-wise weighting scheme (e.g., Theobald and Wuttke 2006), as resistant-fit algorithms can use different subsets of points to locate each configuration in shape space. Resistant-fit is thus an exploratory technique easier to justify as a method for dealing with outliers (e.g., Dryden and Walker 1999) than as an estimation procedure. Resistant-fit did not perform well in a simulation of more a complex case with many degrees of variation in allometry (Walker 2000).

Although we are confident that these subset alignment techniques will prove useful in a variety of data sets, we are well aware that they may not always be appropriate, for example, when the form studied has a complex relationship with size, as in the simulations represented in Fig. 6. Ultimately, what we wish to do is to infer and test a biological model of the underlying transformations that explain the pattern of variation observed in the data (e.g., Zelditch et al. 1990; Richtsmeier et al. 2005; Klingenberg 2009). This process has been attempted surprisingly rarely, in part because practitioners are all too aware of the dependence of inferred variation on the assumptions used to superimpose the data sets.

Finally, we want to emphasize that, while we have confined our analyses and interpretations to allometric variation in shape, the kinds of approaches we take here are applicable to other patterns of variation in shape. Even in the absence of size variation or allometry, biological questions often invite the interpretation of local variation, and the formulation and testing of hypotheses about that variation. The current dismissal by many leading theorists in geometric morphometrics of attempts to infer patterns of variation in shape impoverishes the field by excluding many compelling questions from consideration. As the widespread utilization of visualizations in geometric morphometric studies demonstrates, the intuition is widespread that the patterns of covariance revealed really are meaningful, at least as an exploratory technique. If that intuition is taken seriously analyses should be developed that take advantage of the real information that is present, even if it is confounded to some extent by the superimposition algorithm used. We hope that our modest proposals for alternative approaches to superimposition will help spur full consideration of this deep and important problem.

Computing can be obtained from http://nihroadmap.nih.gov/bioinformatics.

References

Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: Ten years of progress following the 'revolution'. *The Italian Journal of Zoology, 71*(9), 5–16. doi:10.1080/11250000409356545.

Akca, M.D. (2003). Generalized procrustes analysis and its applications in photogrammetry. Available at: http://e-collection.ethbib.ethz.ch/ecol-pool/bericht/bericht_363.pdf.

Bookstein, F. L. (1986). Size and shape spaces for landmark data in two dimensions. *Statistical Science, 1*, 181–242. doi:10.1214/ss/1177013696.

Bookstein, F. L. (1991). *Morphometric tools for landmark data: Geometry and biology*. Cambridge: Cambridge University Press.

Bookstein, F. L. (1996). Biometrics, biomathematics, and the morphometric synthesis. *Bulletin of Mathematical Biology, 58*, 313–365. doi:10.1007/BF02458311.

Buckley, P. F., Dean, D., Bookstein, F. L., Friedman, L., Kwon, D., Lewin, J. S., et al. (1999). Three-dimensional magnetic resonance-based morphometrics and ventricular dysmorphology in schizophrenia. *Biological Psychiatry, 45*(1), 62–67. doi:10.1016/S0006-3223(98)00067-5.

de Celis, J. F. (2003). Pattern formation in the *Drosophila* wing: The development of the veins. *BioEssays, 25*, 443–451. doi:10.1002/bies.10258.

Dickinson, M. H., Lehmann, F., & Sane, S. P. (1999). Wing rotation and the aerodynamic basis of insect flight. *Science, 284*, 1954–1960. doi:10.1126/science.284.5422.1954.

Dryden, I. L., & Mardia, K. V. (1998). *Statistical shape analysis*. Chichester, U.K.: John Wiley and Sons.

Dryden, I. L., & Walker, G. (1999). Highly resistant regression and object matching. *Biometrics, 55*, 820–825. doi:10.1111/j.0006-341X.1999.00820.x.

Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society: Series B, 53*(2), 285–339.

Goodall, C. (1995). Procrustes methods in the statistical analysis of shape revisited. In K. V. Mardia & C. A. Gill (Eds.), *Current issues in statistical shape analysis* (pp. 18–33). Leeds: Leeds University Press.

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika, 40*, 33–51. doi:10.1007/BF02291478.

Held, L. I. (2002). *Imaginal discs: The genetic and cellular logic of pattern formation*. Cambridge: Cambridge University Press.

Houle, D., Mezey, J., Galpern, P., & Carter, A. (2003). Automated measurement of drosophila wings. *BMC Evolutionary Biology, 3*, 25. doi:10.1186/1471-2148-3-25.

Klingenberg, C.P. (2009). Morphometric integration and modularity in configurations of landmarks: Tools for evaluating a-priori hypotheses. *Evolution & Development, 11,* (in press).

Lele, S. (1993). Euclidean distance matrix analysis (EDMA): Estimation of mean form and mean form difference. *Mathematical Geology, 25*(5), 573–602. doi:10.1007/BF00890247.

Lele, S. R., & McCulloch, C. E. (2002). Invariance, identifiability, and morphometrics. *Journal of the American Statistical Association, 97*(459), 796–806. doi:10.1198/016214502388618609.

Lele, S. R., & Richtsmeier, J. T. (2001). *An invariant approach to statistical analysis of shapes*. London, U.K: Chapman and Hall–CRC press.

Mezey, J. G., & Houle, D. (2005). The dimensionality of genetic variation for wing shape in *Drosophila melanogaster*. *Evolution; International Journal of Organic Evolution, 59*(5), 1027–1038.

Mosimann, J. E. (1970). Size allometry: Size and shape variables with characterizations of lognormal and generalized gamma distributions. *Journal of the American Statistical Association, 65*(330), 930–945. doi:10.2307/2284599.

Richtsmeier, J. T., Lele, S. R., & Cole, T. I. (2005). Landmark morphometrics and the analysis of variation. In B. Hallgrímsson & B. K. Hall (Eds.), *Variation: A central concept in biology* (pp. 49–68). New York: Academic Press.

Rohlf, F. J. (2003). tpsRegr, version 1.28–1.30. Department of ecology and evolution, Stony brook, NY: State University of New York.

Rohlf, F. J., & Bookstein, F. L. (2003). Computing the uniform component of shape variation. *Systematic Biology, 52*(1), 66–69. doi:10.1080/10635150390132759.

Rohlf, F. J., & Marcus, L. F. (1993). A revolution in morphometrics. *Trends in Ecology & Evolution, 8*, 129–132. doi:10.1016/0169-5347(93)90024-J.

Rohlf, F. J., & Slice, D. (1990). Extensions of the procrustes method for the optimal superimposition of landmarks. *Systematic Zoology, 39*(1), 40–59. doi:10.2307/2992207.

Siegel, A. F., & Benson, R. H. (1982). A robust comparison of biological shapes. *Biometrics, 38*(2), 341–350. doi:10.2307/2530448.

Slice, D. E. (1998). Morpheus et al.: software for morphometric research, version revision 01-30-98. Department of Ecology and Evolution, Stony Brook, NY: State University of New York.

Slice, D. (2001). Landmark coordinates aligned by procrustes analysis do not lie in Kendall's shape space. *Systematic Biology, 50*, 141–149. doi:10.1080/10635150119110.

Sun Microsystems Inc (1992–2006). Java(tm) development kit, version 1.1.8_005. Santa Clara, CA: Sun Microsystems, Inc.

Theobald, D. L., & Wuttke, D. S. (2006). Empirical Bayes hierarchical models for regularizing maximum likelihood estimation in the matrix Gaussian procrustes problem. *Proceedings of the National Academy of Sciences of the United States of America, 103*(49), 18521–18527. doi:10.1073/pnas.0508445103.

van der Linde, K. & Houle D. (2004–2008). Wings. Tallahassee: Florida State University. Available at: http://www.kimvdlinde.com/morphometrics.

van der Linde, K. (2005–2008). Geometrics package. Tallahassee: Florida State University. Available at: http://www.kimvdlinde.com/morphometrics.

Verboon, P., & Gabriel, K. R. (1995). Generalized procrustes analysis with iterative weighting to achieve resistance. *The British Journal of Mathematical and Statistical Psychology, 48*, 57–73.

Walker, J. A. (2000). Ability of geometric morphometric methods to estimate a known covariance matrix. *Systematic Biology, 49*(4), 686–696. doi:10.1080/106351500750049770.

Zelditch, M. L., Straney, D. O., Swiderski, D. L., & Carmichael, A. C. (1990). Variation in developmental constraints in *Sigmodon*. *Evolution; International Journal of Organic Evolution, 44*, 1738–1748. doi:10.2307/2409503.

Zelditch, M. L., Swiderski, D. L., Sheets, H. D., & Fink, W. L. (2004). *Geometric morphometrics for biologists: A primer*. Amsterdam: Elsevier.