Testing and Quantifying Phylogenetic Signals and Homoplasy in Morphometric Data

CHRISTIAN PETER KLINGENBERG1,* AND NELLY A. GIDASZEWSKI1,2

1Faculty of Life Sciences, University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK; and
2Present address: Département Systématique et Evolution, Muséum National d’Histoire Naturelle, UMR CNRS 7205, 45 rue Buffon, 75005 Paris, France; E-mail: nellygid@yahoo.fr.

Received 13 April 2009; reviews returned 9 July 2009; accepted 9 November 2009

Abstract.—The relationship between morphometrics and phylogenetic analysis has long been controversial. Here we propose an approach that is based on mapping morphometric traits onto phylogenies derived from other data and thus avoids the pitfalls encountered by previous studies. This method treats shape as a single, multidimensional character. We propose a test for the presence of a phylogenetic signal in morphometric data, which simulates the null hypothesis of the complete absence of phylogenetic structure by permutation of the shape data among the terminal taxa. We also propose 2 measures of the fit of morphometric data to the phylogeny that are direct extensions of the consistency index and retention index used in traditional cladistics. We apply these methods to a small study of the evolution of wing shape in the Drosophila melanogaster subgroup, for which a very strongly supported phylogeny is available. This case study reveals a significant phylogenetic signal and a relatively low degree of homoplasy. Despite the low homoplasy, the shortest tree computed from landmark data on wing shape is inconsistent with the well-supported phylogenetic tree from molecular data, underscoring that morphometric data may not provide reliable information for inferring phylogeny. [Drosophilidae; geometric morphometrics; homoplasy; phylogeny; Procrustes superposition; shape; squared-change parsimony.]

Understanding the evolution of organismal form is a primary goal of comparative biology, and inferring the phylogenetic history of shape change is therefore a central concern. Nevertheless, the question of how morphometric data should be used in the context of phylogenetic analyses has long been hotly debated (e.g., Felsenstein 1988, 2002; Zelditch et al. 1995; Naylor 1996; Collard and Wood 2000; Monteiro 2000; Cannon and Manos 2001; Polly 2001; MacLeod 2002; Rohlf 2002; Lockwood et al. 2004; Caumul and Polly 2005; Lycett and Collard 2005; Cardini and Elton 2008; González-José et al. 2008). The disagreement primarily concerns the question whether morphometric data can be used to estimate phylogeny. In contrast, there is no major controversy concerning analyses that rely on the mapping of shape variation onto an independent and well-supported phylogeny.

Detecting and testing phylogenetic “signals” or “structure” in morphometric data is a part of this type of analysis (Cole et al. 2002; MacLeod 2002; Cardini and Elton 2008) and extends analyses that have long been used with conventional characters (e.g., Kitching et al. 1998). Phylogenetic signals in some aspect of the phenotype can be defined as the degree to which phylogenetic relatedness among taxa is associated with their phenotypic similarity (e.g., Blomberg et al. 2003; Cardini and Elton 2008). In other words, a strong phylogenetic signal exists if closely related taxa are phenotypically more similar than taxa that are phylogenetically more remote. Phylogenetic signals can be sought in all aspects of the phenotype, including ecological and physiological traits. In this paper, however, we will focus on morphometric variation.

Clear phylogenetic signals are expected for traits that are undergoing divergent evolution. In a representation of such evolution in a phenotypic space, different lineages of a clade evolve from the ancestral position in different directions. Phylogenetic signals are diluted by processes leading to homoplasy, such as reversals, parallel evolution, and convergence. Therefore, the degree of homoplasy in morphometric data is expected to be negatively associated with the strength of phylogenetic signal. Previous studies have assessed the strength of phylogenetic signals by quantifying the congruence between phylogenetic trees estimated by various methods and phenetic trees constructed from the morphometric data (e.g., Loy et al. 1993; David and Laurin 1996; Cannon and Manos 2001; Cole et al. 2002; Houle et al. 2003; Caumul and Polly 2005; Leinonen et al. 2006; Neustupa and Škaloud 2007; Young 2008). An alternative to the use of congruence between trees is the mapping of the morphometric traits onto the phylogeny by squared-change parsimony or similar methods (Huey and Bennett 1987; Maddison 1991; McArdle and Rodrigo 1994; Rohlf 2001). Using these or similar methods, phylogenetic trees can be projected into a morphometric space to visualize the phylogenetic history of morphometric traits (Klingenberg and Eckau 1996; Rohlf 2002; Miller and Venable 2003; Nicola et al. 2003; Linde et al. 2004; Stayton 2005; Macholán 2006; Schlick-Steiner et al. 2006; Sidlauskas 2008). In combination with ordination techniques such as principal components or canonical variate analysis, this method can be used to investigate the evolution of shape just as it has long been done for traditional characters (e.g., Brooks and McLennan 1991).

Here we present a new procedure for testing the existence of a phylogenetic signal in morphometric data, which is based on mapping morphometric data onto a known phylogeny and extends a test described...
by Laurin (2004) for scalar traits. The proposed test uses a permutation approach to simulate the null hypothesis of the complete absence of phylogenetic structure by randomly reassigning the phenotypic data to the terminal nodes of the phylogeny. Detecting a phylogenetic signal is closely related to measuring the fit of the phenotypic data to a given phylogeny for quantifying the homoplasy in the data. In traditional cladistic analysis, this has been done with measures such as the consistency index or retention index (e.g., Kitching et al. 1998, chapter 5). We define analogous measures for morphometric data and demonstrate how these indices can be computed in the context of geometric morphometrics. This approach is applicable to a wide range of morphometric data. We illustrate the new methods with an example on the evolution of wing shape in the Drosophila melanogaster subgroup, for which the phylogeny has been studied extensively (e.g., Lachaise et al. 2003). We characterize wing shape with landmark data and Procrustes methods (Dryden and Mardia 1998). Our analyses demonstrate that these data on wing shape, despite a low level of homoplasy, fail to estimate the phylogeny of this group correctly.

**Geometric Morphometrics and “Characters”**

A central issue in the controversy over the use of geometric morphometrics in the context of phylogeny is the question whether shape can be used as a character or a set of separate characters and how a partitioning into separate characters should be done (e.g., Bookstein 1994; Zelditch et al. 1995; Monteiro 2000; MacLeod 2002). Some methods of phylogenetic inference from quantitative traits, such as neighbor-joining and other clustering methods, are sidestepping this issue because they use information on distances among taxa rather than character states (Swofford et al. 1996; Felsenstein 2004). Similarly, likelihood-based methods can be reduced to considerations of distances (Felsenstein 1973), so that the choice of a particular set of shape variables will not affect the result as long as the complete information about the relative locations of taxa in the multidimensional shape space is used. For these approaches, the definition of characters is not a central issue and the notion of characters might even be abandoned altogether. In contrast, the traditional methods of cladistics based on Wagner parsimony and similar methods assume that characters are separate from one another, so that they can serve independent sources of phylogenetic information (Kitching et al. 1998). For this approach, it is therefore of central importance how the multidimensional shape variation is partitioned into separate characters and how character states are defined.

**Shape as a Set of Multiple Characters**

A possibility of combining morphometrics and phylogenetic analysis is to use the shape variables as cladistic characters. One proposed solution was to use partial warp scores from landmark data in this manner (e.g., Fink and Zelditch 1995; Zelditch et al. 1995, 2000). This method caused a considerable controversy (e.g., Bookstein 1994, 2002; Naylor 1996; Adams and Rosenberg 1998; Rohlf 1998; Monteiro 2000; MacLeod 2002) and later was effectively retracted by the original authors (Zelditch et al. 2004, chapter 14). A central element of the criticism of the method was that the choice of shape descriptors, the partial warps as opposed to some alternative coordinate system, was largely arbitrary.

Extracting characters from a multidimensional shape space can impose an apparent ordering of taxa, which depends on the choice of a variable in the multidimensional space. As a simple example, consider 3 shapes A, B, and C that form a triangle in a 2-dimensional shape space. If shape variables are defined by the projection of the 3 points onto a line in the shape space, projections in different directions can yield the sequences ABC, ACB, and BAC (or in the reverse, BCA, BCA, and CAB). Each shape variable thus provides an ordering, and any ordering of the shapes is possible, depending on the choice of the direction used for the projection. Likewise, different projections can generate all 3 possible contrasts of 2 shapes against the third, which may be interpreted as binary characters (A and B versus C, A and C versus B, B and C versus A). The choice of the shape variables induces the ordering or grouping of the values for the taxa and thus determines the results of the phylogenetic analyses in which these variables are used as characters. Monteiro (2000) discussed the implications of these choices for phylogenetic studies in more detail.

A different approach is to subdivide the shape into smaller parts and to derive shape variables as characters from them (MacLeod 2002; González-José et al. 2008). For instance, MacLeod (2002) subdivided the body shape of fish into portions corresponding to named anatomical regions (e.g., “mouth region,” “abdomen,” and “tail”) and coded shape differences in these regions as separate characters in the subsequent phylogenetic analysis. González-José et al. (2008) separated the hominid skull into 4 parts and used the first few principal components of each as continuous characters. In this approach, the choice of the parts may have a considerable influence on the results, in addition to the shape variation in the data.

Most studies that extracted sets of multiple characters from shape data have used Wagner parsimony or related approaches (Kluge and Farris 1969; Kitching et al. 1998). In this context, characters are the features of organisms about which separate hypotheses of homology are made (e.g., Kitching et al. 1998, chapter 2). To compare different phylogenetic hypotheses, these methods compute the total change on the respective trees by mapping the characters onto the tree and adding up changes of individual characters. To avoid counting multiple character changes for single events where multiple characters change together due to a common underlying cause, these methods assume that characters are independent of each other (i.e., that there are no such common underlying causes).
Some of the controversy about the use of morphometric data in phylogenetics relates to this assumption (Rohlf 1998; Monteiro 2000; Bookstein 2002). Shape variables extracted from a morphological structure jointly make up the shape as a whole and normally cannot be assumed to be independent of each other. Different shape variables may carry the same information, for instance, if size has a major influence on all aspects of a shape through allometry or if other factors promote integration of traits (e.g., Klingenberg 2008b). The shapes of different body parts are often found to be highly integrated with each other (e.g., Klingenberg et al. 2001; Young and Hallgrimsson 2005), and even parts that behave as distinct modules may not be fully independent but show some degree of integration across a larger structure or throughout the whole organism (Klingenberg et al. 2003; Klingenberg 2009; Laffont et al. 2009). Such integration among parts makes it questionable whether shape variables derived from them can be used as separate characters. Even using uncorrelated shape variables, as they can be obtained from a principal component analysis of terminal taxa (e.g., MacLeod 2002; González-José et al. 2008), does not necessarily solve the problem of character interdependence because the covariation among terminal taxa is not generally the same as the covariation among evolutionary changes in shape variables (Felsenstein 1985, 2002, 2004).

**Shape as a Single Multidimensional Character**

An alternative to these methods of subdividing shape into separate characters is to treat shape as just one character, albeit a complex one that is inherently multidimensional (another option is to drop the notion of characters altogether and to focus on evolutionary change as shifts in phenotypic space). Treating shape as a single character involves a generalization of the continuous characters familiar from conventional studies, where they represent a measurement of a scalar quantity such as a length or weight. Biological shapes can change continuously by arbitrarily small local displacements of parts or overall deformations, but the difference to scalar characters such as length measurements is that shapes can differ in many different ways. For example, a fly wing can become narrower and more elongated or there can be shifts of individual veins relative to one another. Accordingly, the difference between 2 shapes has not only a magnitude but also a direction in the shape space (e.g., Klingenberg and Monteiro 2005). Therefore, phylogenetic analyses of shape need to adapt the methods normally used for continuous characters by applying multivariate generalizations.

With shape as a single character, entire shapes are character states. In the context of parsimony analysis, the cost matrix used in the generalized parsimony approach (Swofford et al. 1996) can be generalized using a cost function. Just as for scalar-valued continuous characters, a function \( f(x,y) \) characterizes the cost of the change from state \( x \) to state \( y \) (where \( x \) and \( y \) are vectors representing the respective shapes). In other words, the function \( f(x,y) \) quantifies the minimum amount of evolutionary change required from state \( x \) to state \( y \). Maximum parsimony maps a set of shapes for terminal nodes onto a phylogenetic tree by finding the set of shapes for the internal nodes that minimizes the sum of the values of the cost function over all branches of the tree (this sum is usually called the tree length). This overall approach is the same for all types of parsimony; what differs between them is the cost function and, often, the algorithms used to minimize the tree length. Discrete characters differ in that the corresponding cost functions are just defined for a finite set of character states (e.g., 0 and 1) and therefore can be represented as cost matrices (Swofford et al. 1996). Different variants of parsimony such as Wagner, Dollo, or Fitch parsimony differ in the values of the cost function for specific transitions from one character state to another (Swofford et al. 1996; Kitching et al. 1998).

The most widespread parsimony method for multivariate phenotypes is squared-change parsimony (Huey and Bennett 1987; Maddison 1991; McArdle and Rodrigo 1994; Rohlf 2001, 2002). It has been used extensively for displaying evolutionary change of morphometric traits (e.g., Klingenberg and Ékau 1996; Rohlf 2002; Miller and Venable 2003; Nicola et al. 2003; Linde et al. 2004; Macholán 2006; McPeek et al. 2008; Scholtz et al. 2008; Sidlauskas 2008; Astúa 2009). For this method, the cost function \( f(x,y) \) is the squared Euclidean distance between \( x \) and \( y \) in the feature space of the morphometric data. The tree length is computed as the sum of squared changes, summed over all branches and all the coordinates. For the context of geometric morphometrics with landmark data, this means that the cost function \( f(x,y) \) is the squared Procrustes distance between shapes \( x \) and \( y \) (the tangent space approximation is appropriate even for analyses at a large phylogenetic scale, e.g., for skull shapes across all mammals; Marcus et al. 2000). For other data types, squared Euclidean distances can be computed in the respective spaces of shape descriptors. Accordingly, squared-change parsimony has been applied to semilandmarks (Macholán 2006), other types of outline data (Linde et al. 2004; Scholtz et al. 2008), and 3-dimensional surface data (McPeek et al. 2008). Squared-change parsimony has a number of statistical advantages: it provides unambiguous point estimates of the shapes at internal nodes of the phylogeny, is invariant under rotation of the coordinate system, and relates directly to the null model of Brownian motion in the feature space (Maddison 1991; Rohlf 2001, 2002). For these reasons, we will use it extensively in this study.

A possible alternative is the minimum evolution method of Cavalli-Sforza and Edwards (1967; see also Thompson 1973), which minimizes the sum of the Euclidean distances along all the branches of the phylogeny (it may be preferable to call this method “Euclidean parsimony” to avoid confusion with the now
more widespread “minimum evolution” method of Rzhetsky and Nei 1992). This method also provides unambiguous point estimates for phenotypes at internal nodes and is invariant under rotation of the coordinate system (for discussion of properties and algorithms, see Thompson 1973; Smith 1992). In the context of shape studies with landmark data, the cost function for this method is the Procrustes distance. This method has been used only rarely in evolutionary biology.

Using shape as a single multivariate character is a direct generalization of the approaches developed for scalar continuous characters, but there are differences that are due to the special restrictions of scalar traits to 1 dimension. The observed values of a scalar character can always be arranged on a single line, which can be interpreted as a segment of the number line covering the range of the values (Fig. 1a). This segment of the number line is also the character state tree for this set of values (under Wagner and squared-change parsimony, among others). In contrast, multidimensional characters generally cannot be arranged in a single dimension, and the character state tree has a branching structure (Fig. 1b, under squared-change parsimony; Wagner parsimony is not applicable). We will use this concept of shape as a single, multidimensional character in the remainder of this paper to test and quantify phylogenetic signals.

Figure 1. Scalar and multidimensional characters. a) A scalar continuous character. Its character states (dots) are real values, and the character states can therefore be arranged along the number line between the 2 most extreme values (the range of the character). The character state tree for such a character is not branched but is the same segment of the number line. The minimum amount of evolutionary change is the length of this tree (e.g., for Wagner parsimony, it is the range of the values). b) A multidimensional continuous character. In general, character states cannot be arranged along a single line, and projections in different directions result in different orderings of character states. Usually, the character state tree is branched, as shown here for an example where 4 character states form a square (there is an equivalent version of this tree, for which the central branch is vertical and not horizontal).

**Permutation Test of Phylogenetic Signal**

A strong phylogenetic signal means that the shapes of closely related species tend to be more similar to each other than the shapes of more distantly related species. In other words, closely related species should occupy the same portion of the morphometric space, whereas more distantly related species should be found in distinct and possibly distant locations. A way to analyze this link between relatedness and morphometric distances is to map the morphometric data onto the phylogenetic tree, which can be visualized directly as a projection of the phylogenetic tree into the morphometric space (Klingenberg and Ekau 1996; Rohlf 2002; Miller and Venable 2003; Nicola et al. 2003; Linde et al. 2004; Stayton 2005; Macholán 2006; Sidlauskas 2008). This involves an explicit reconstruction of the morphometric data for the ancestral nodes of the tree, usually with the method of squared-change parsimony (Maddison 1991; Rohlf 2001, 2002). In the presence of a strong phylogenetic signal in the data, closely related species tend to be near each other in morphometric space and, as a result, the average amount of shape change along the branches of the tree is relatively small. In contrast, data lacking a phylogenetic signal tend to produce greater shape changes on the branches of the phylogeny because closely related species are expected to be just as distant from each other as remotely related species. Overall, the expected amount of change on the entire tree is smaller if there is a strong phylogenetic signal than in the absence of a phylogenetic signal.

Because the strength of the phylogenetic signal is a property of the entire phylogeny and the morphometric data set jointly, a summary of the amounts of shape change over the whole tree is needed. A simple summary is the sum of squared changes along all branches, which is the criterion that is minimized in squared-change parsimony (Maddison 1991). If there is a strong phylogenetic signal, related species will be close to each other, and the sum of squared changes is thus expected to be relatively small. In the absence of a phylogenetic signal, however, related species will not always be close to each other, which will result in greater changes along the branches of the phylogeny and thus the sum of all squared changes is expected to be higher. The sums of squared changes need to be evaluated in relation to the particular phylogeny and morphometric data under study.

To assess an observed amount of morphometric change statistically, the null hypothesis of the total absence of a phylogenetic signal can be simulated by a permutation procedure. If there is no effect of phylogeny, then random swapping the morphometric values among the tips of the phylogenetic tree should not affect the expected total of morphometric change along the branches of the tree because a different assignment of the morphological data to the branch tips is equally likely to produce a greater or lesser amount of total change than is observed for the original data. In contrast, if there is a strong phylogenetic signal,
then exchanging the morphometric values among the tips of the tree will most likely result in a longer tree than the one originally obtained. The null hypothesis of no phylogenetic signal can therefore be simulated by repeatedly and randomly exchanging the morphometric values among the terminal taxa of the phylogeny, mapping the permuted morphometric data onto the phylogeny, and calculating the total tree length in units of morphometric distance. The empirical $P$ value for the test is the proportion of permuted data sets in which the sum of squared changes is shorter or equal to the value obtained for the original data.

This test is a direct multivariate extension of the “random taxon reshuffling” test proposed by Laurin (2004, p. 599 f.). Laurin designed that test to ascertain a phylogenetic signal in a scalar quantitative trait such as body size, but it is easily extended in the new context of the multivariate analysis of shape. The null hypothesis of a complete absence of a phylogenetic signal is closely related to those used in other permutation tests in contexts such as regression analysis (e.g., Good 2000), and it is the proper null hypothesis for a test for the presence or absence of a phylogenetic signal.

### Quantifying Homoplasy in Shape Data

Whereas the permutation test evaluates the presence or absence of a phylogenetic signal, it does not quantify its strength. For scalar characters, the strength of a phylogenetic signal is usually assessed by computing measures such as the consistency index (Kluge and Farris 1969) and the retention index (Archie 1989; Farris 1989; for a review of these indices, see Kitching et al. 1998, chapter 5). Although these indices have mostly been computed for discrete characters, they can also be used with continuous characters: Kluge and Farris (1969) explicitly defined the consistency index to be applicable to continuous as well as discrete characters, and the definitions of statistics such as the retention index are equally applicable to continuous data (Ackerly and Donoghue 1998). To assess how well the morphometric data fit a phylogenetic tree or how much homoplasy is present, we offer a multivariate generalization of both indices in the context of squared-change parsimony.

### Shape Consistency Index

For a scalar character, the consistency index is computed as $m/s$, where $m$ is the minimal amount of change needed to map the character onto any phylogeny and $s$ is the amount of change needed for the tree under investigation (Kluge and Farris 1969; Farris 1989; Kitching et al. 1998, p. 95). Both $m$ and $s$ depend on the distribution of character states and on the optimization criterion used for mapping characters onto phylogenies. The value of $s$ is obtained by mapping the character onto the tree. For a discrete character and one of the conventional parsimony criteria (e.g., Wagner, Fitch, or Dollo parsimony; Kitching et al. 1998, chapter 4), the value of $m$ is the number of character states minus one. For a scalar continuous character and Wagner parsimony, the value of $m$ is the range of the character (Kluge and Farris 1969, p. 8). Generally, $m$ can be computed as the minimum length of a tree that connects all the character states under the particular type of parsimony (Fig. 1a).

The consistency index takes positive values less than or equal to 1.0. High values of the consistency index indicate a low degree of homoplasy in the data set. Low values indicate large amounts of homoplasy, but the index never reaches zero even with extremely high homoplasy (see Kitching et al. 1998, p. 95–97).

The computation of the consistency index can be extended for shape as a multidimensional character to define a “shape consistency index,” by computing $m$ as the minimal length of any tree fitting the set of shapes of the terminal taxa (Fig. 1b) and $s$ as the length for the tree with the topology of the phylogeny of interest. For shape data, the tree length $s$, measured as the sum of the squared shape changes along all branches, can be obtained by fitting the shape data to the tree with the method of squared-change parsimony (Maddison 1991; Rohlf 2002). Finding $m$ for a given set of taxa is substantially more difficult because it requires finding the tree that connects the shapes of the taxa and that has the minimal length. This has been known in computer science as the Steiner tree problem and has long been an area of active research (e.g., Prömel and Steger 2002). The link between phylogenies and the Steiner tree problem was already made by Cavalli-Sforza and Edwards (1967, p. 563) and in detail by Thompson (1973). Effective algorithms for finding Euclidean Steiner trees for multidimensional data have been developed since then (Prömel and Steger 2002). At least for small numbers of taxa, the minimal-length tree can be found with the branch-and-bound algorithm proposed by Smith (1992), which can be modified for squared-change parsimony (Fig. 2a). For larger numbers of taxa, heuristic algorithms need to be used (Prömel and Steger 2002).

These minimum-length trees should not be confused with minimum spanning trees, which have sometimes been used in similar contexts (Snath and Sokal 1973; Loy et al. 1993; Rohlf et al. 1996; Cardini and O’Higgins 2004; Claude et al. 2004; Monteiro and dos Reis 2005; Cardini et al. 2007). Minimum spanning trees are defined as trees in which the branches directly connect the observed taxa without any additional nodes corresponding to hypothetical ancestors (Fig. 2b). Because they do not reflect ancestry relations among taxa, minimum spanning trees cannot represent the phylogenetic changes of shape. Also, because the method precludes internal nodes, minimum spanning trees are usually longer than minimal Steiner trees for the same distance measure (e.g., Fig. 2a,b).

Like the consistency index for scalar characters, the shape consistency index has a positive value up to a maximum of 1.0, with high values indicating low amounts of homoplasy. Because we consider shape as a single, multidimensional character, the shape consistency index corresponds to the character consistency...
**Figure 2.** A square of 4 taxa in a simplified shape space and 3 possible trees that can be constructed for the taxa. The coordinates of the 4 taxa and of internal nodes are indicated as \((x, y)\) values. 

(a) One of the 2 trees with the minimal sum of squared branch lengths. The other minimal tree is rotated by 90° so that its internal branch is in vertical orientation. For a square with sides of length 1.0, the length of these 2 trees is 1.5 (sum of squared branch lengths).

(b) A minimum spanning tree. By definition, a minimum spanning tree links the taxa directly. It therefore has no ancestral nodes and is unsuitable as a representation of a phylogenetic tree.

(c) The tree for the star phylogeny (completely unresolved bush). All branches emanate from a single internal node at the centroid of the 4 taxa. The length of this tree is 2.0, which is the maximum tree length reconstructed by squared-change parsimony for any phylogenetic tree. The tree topology that yields the maximal tree length is the star phylogeny or completely unresolved bush, in which all the shape changes among taxa are mapped to separate branches that go from the single shared ancestor to each of the taxa (Fig. 2c). Because none of the similarity among taxa is due to shared ancestry in a star phylogeny, this topology maximizes the amount of homoplasy. Under squared-change parsimony, the location of the ancestral node of the star phylogeny can be calculated as the mean of all the shapes in the set of terminal taxa, and the tree length \(g\) is the sum of squared coordinate differences of the taxa from this mean shape.

The retention index can take values from zero to one. It can be interpreted as a measure of the degree of synapomorphy in the data set (Kitching et al. 1998, p. 97–99). The terms \((g - s)\) and \((g - m)\) can be understood as the degrees to which the tree of interest and the shortest possible tree for the data improve over the “worst-case scenario” of the star phylogeny with a single internal node. In this scenario, there is no synapomorphy because all variation among taxa is from changes independently accrued in the different lineages, and homoplasy is maximal. The improvement over the star phylogeny is therefore interpretable as a measure of the degree of synapomorphy in the data. The retention index can therefore be interpreted as a measure of synapomorphy in the tree of interest relative to the range between the extreme cases of the shortest and longest possible trees.

**Computing the Indices**

The computation of these measures of homoplasy has implications for the choice of methods for mapping shape data onto phylogenies. Both the consistency index and the retention index are based on the comparison of the actual tree to the minimal- and maximal-length trees for the set of character states. Because these are character state trees, both may have fundamentally different tree topologies from the evolutionary tree of the taxa under study, and there is thus no clear correspondence between the branches of the different trees. Therefore, there is no possibility for weighting changes by the inverse of the branch lengths in a specified tree. Accordingly, the indices must be computed with unweighted squared-change parsimony, that is, with all branch lengths set to equal values (Maddison 1991; McArdle and Rodrigo 1994; Rohlf 2001).

Similar considerations also lead to the conclusion that unrooted rather than rooted trees are preferable for computing tree lengths in this context. Because fundamentally different tree topologies are used to compute the
actual, minimal, and maximal tree lengths, it is not clear how they should be rooted (this particularly applies to the minimum tree). Therefore, unrooted versions of all 3 trees should be used. For squared-change parsimony, this choice does make a difference for the overall tree length (Maddison 1991). In a rooted tree, 2 separate branches connect the root to the adjacent nodes (Fig. 3a). In the unrooted version of the same tree, however, these 2 branches are united into a single branch (Fig. 3b). In the rooted tree, the total trait change between the nodes adjoining to the root is divided into 2 smaller differences that are distributed over the 2 branches. The sum of the squares of these 2 smaller differences is less than the square of the total difference, which is associated with the single branch in the unrooted tree (if one or both of the adjoining nodes are internal, the reconstructed values themselves will not be identical). Therefore, the lengths of unrooted trees computed by squared-change parsimony will tend to be longer than the lengths of rooted trees obtained for the same data. In order to obtain an optimal comparison, unrooted trees should be used for the actual, minimal, and maximal trees for the shape consistency and retention indices. Algorithms for computing tree lengths by squared-change parsimony normally use rooted trees; the lengths for unrooted trees can be obtained by re-rooting them at an internal node, resulting in a basal trichotomy (Fig. 2c; Maddison 1991; Rohlf 2001).

**Materials and Methods**

We illustrate the methods introduced in this paper with an example that focuses on the evolution of wing shape in flies. We use landmarks to characterize the wing shapes in this example (Fig. 4). Although additional features, such as the contours of the outline and wing veins, would provide additional information that can be incorporated with appropriate methods (e.g., Polly 2008), these methods involve assumptions about homology of points that cannot be assessed with morphological evidence, introducing other potential problems (Klingenberg 2008c). We emphasize, however, that the methods presented in this paper can be used in conjunction with those analyses as well because squared-change parsimony can be used in those contexts, as has been demonstrated for semilandmarks (Macholán 2006), outlines (Linde et al. 2004; Scholtz et al. 2008), and 3-dimensional surfaces (McPeek et al. 2008).

**Species Used and Morphometric Data**

We apply the methods introduced above in a case study of flies from all 9 known species of the *D. melanogaster* subgroup: *D. melanogaster*, *D. simulans*, *D. mauritiana*, *D. sechellia*, *D. yakuba*, *D. santomea*, *D. teissieri*, *D. orea*, and *D. erecta*. The flies were obtained from long-established laboratory cultures. As far as possible, 50 females and 50 males of each species were used, but the final sample sizes ranged from 34 to 53 flies per species and sex. The right wings were mounted on microscope slides, digital images were taken, and a set of 15 landmarks was digitized for each wing (Fig. 4). The landmarks were chosen to cover the wing blade as completely as possible, while ensuring that they could be located precisely in all specimens and were unambiguously homologous across all 9 species.

Shape information was extracted from the landmark configurations of the combined samples of all species by generalized Procrustes superimposition and projection to the tangent space (e.g., Dryden and Mardia 1998). The mean shapes for each species were then computed by averaging wing shapes of all males and females (sex dimorphism is analyzed in Gidaszewski et al. 2009). To display the arrangement of the data points in the shape tangent space, we used principal component analysis of the covariance matrix of shape tangent coordinates, computed from the 9 species means. A plot of the first 2 principal components is an optimal representation in 2 dimensions because it is the plot that shows the maximal amount of the variation among species means (e.g., Jolliffe 2002).

**Phylogenetic Trees**

We used 2 published phylogenies of the *D. melanogaster* subgroup that included all 9 species considered here and a composite phylogeny reconstructed from different published sources (Fig. 5). The phylogenetic trees from single sources were a tree for the *period* gene (Lachaise 2010). Klingenberg and Gidaszewski—Phylogenetic Signal in Shape Data
et al. 2000, figure 4b) and one for the Amyrel gene (Cariou et al. 2001, figure 4). The combined tree was established from these trees and several other published sources (Lemeunier and Ashburner 1984; Harr et al. 1998; Matsuo 2000; Kastanis et al. 2003; Parsch 2003; Lewis et al. 2005; Drosophila 12 Genomes Consortium 2007). For analyses using weighted squared-change parsimony, we used estimates of divergence times to scale the branch lengths (Russo et al. 1995; Li et al. 1999; Kliman et al. 2000; Llopart et al. 2002; Tamura et al. 2004; McDermott and Kliman 2008).

The most important difference between published trees is the rooting of the *D. melanogaster* subgroup. A number of studies, using a variety of methods and sources of information, yielded phylogenies or estimates of divergence times in which the subgroup consists of 2 main clades, one consisting of *D. melanogaster*, *D. simulans*, *D. mauritiana*, and *D. sechellia* and another consisting of the remaining 5 species (Lemeunier and Ashburner 1984; Matsuo 2000; Arhontaki et al. 2002; Kastanis et al. 2003; Parsch 2003; Tamura et al. 2004; Lewis et al. 2005; Drosophila 12 Genomes Consortium 2007). Other studies present a topology in which *D. orena* and *D. erecta* form a sister clade to the remaining species (Lachaise et al. 1988; Russo et al. 1995, figure 2; Lachaise et al. 2000). Because the first of these 2 possible root locations has overwhelming support from phylogenetic analyses of entire chromosomes and whole genomes (*Drosophila 12 Genomes Consortium 2007*), we used it for the combined tree. For the trees of Lachaise et al. (2000) and Cariou et al. (2001), we used both possibilities of rooting.

A second difference among studies are the relationships among *D. simulans*, *D. mauritiana*, and *D. sechellia*, which are poorly resolved and may represent a trichotomy with independent origins of the 2 island species from *D. simulans* (Kliman et al. 2000; Ting et al. 2000; Parsch 2003; McDermott and Kliman 2008). A later divergence of *D. mauritiana* and *D. simulans*, and therefore a closer relationship between these 2 species, is suggested by microsatellite and sequence data (Harr et al. 1998; Kliman et al. 2000; Harr and Schlötterer 2004). Altogether, the trees in Figure 5 represent both rooting options and the 3 possible arrangements of *D. simulans*, *D. sechellia*, and *D. mauritiana*.

Reconstructing Ancestral Shapes

We used the criterion of squared-change parsimony (both unweighted and weighted by divergence time or molecular change on the respective branches of the tree) for reconstructing the values of the internal nodes of the phylogeny from the shape averages of the terminal taxa (Maddison 1991; McArindle and Rodrigo 1994; Rohlf 2001, 2002). The calculations used the generalized least-squares method (Martins and Hansen 1997; Rohlf 2001) to find values for the internal nodes so that the sum of squared changes along the branches is minimized over the entire phylogeny.

This method can be applied to multidimensional features such as shape, for which it minimizes the sum of squared changes over all variables. This is equivalent to minimizing the squared Euclidean distances along the branches of the phylogeny and is invariant to the effects of rotation (Rohlf 2001). In particular, in the context of geometric morphometrics, minimizing the sum of squared changes in the shape tangent space is equivalent to minimizing the length of the tree as measured in units of squared Procrustes distance and therefore relates directly to the basic metric for shape differences (e.g., Bookstein 1996; Dryden and Mardia 1998). The method of squared-change parsimony has

---

**FIGURE 4.** A wing of *Drosophila melanogaster* with the landmarks used in this study.
Figure 5. Alternative phylogenetic trees for the *Drosophila melanogaster* subgroup. There is a strong overall consensus among these topologies, which differ only in the location of the root and the arrangement of *D. mauritiana*, *D. simulans*, and *D. sechellia*. a) The composite tree compiled from different sources (Lemeunier and Ashburner 1984; Harr et al. 1998; Matsuo 2000; Kastanis et al. 2003; Parsch 2003; Lewis et al. 2005; *Drosophila 12 Genomes Consortium* 2007). b) The tree for the *period* gene (Lachaise et al. 2000), re-rooted to conform to the results of other studies (for discussion, see the text). c) The same tree, rooted as in the original publication (Lachaise et al. 2000, figure 4b). d) The tree for the *amylase-related* gene (Cariou et al. 2001), re-rooted. e) The same tree, rooted as in the original publication (Cariou et al. 2001, figure 4).

been widely used for the explicit mapping of landmark data onto phylogenies (e.g., Rohlf 2002; Nicola et al. 2003; Sidlauskas 2008; Astúa 2009).

**Permutation Test of Phylogenetic Signal**

The null hypothesis of the complete absence of any phylogenetic signal was simulated by permuting the shape means among the species. Because only 9 species were included in the analyses, it was feasible to use a permutation approach based on the complete enumeration of all permutations. For 9 species, the number of possible permutations, $9! = 362,880$, is sufficiently small to be enumerated completely (due to symmetries in the phylogenetic tree, it may be possible to reduce substantially the number of required permutations; however, the algorithm for such tree-dependent permutation would be fairly complex). For studies with more terminal taxa, a large number of random permutations can be used to simulate the null hypothesis (we recommend to use 10,000 or more). The permuted data were mapped onto the phylogeny with both unweighted and weighted squared-change parsimony, and the resulting tree lengths were stored. The empirical $P$ values were obtained as the proportions of the permutations in which the tree lengths were equal to or shorter than those obtained for the original data. The permutation test of phylogenetic signal is implemented in the MorphoJ software (Klingenberg 2008a).

As a demonstration of the approach, we performed the permutation tests for all the tree topologies shown in Figure 5, using both weighted and unweighted squared-change parsimony for each tree. As a result, we are faced with multiple tests. Because the data are the same for all tests and the trees are similar, however, the corresponding tests are largely based on the same information and definitely not independent, as is assumed by methods used to adjust for multiple tests such as the Bonferroni correction. We therefore present the nominal $P$ values.

**Measures of Homoplasy**

To obtain the shape consistency index and shape retention index, it was necessary to compute the minimal and maximal amounts of change that would be reconstructed for any tree topology with squared-change parsimony and the given morphometric data.

**Minimal amount of change.**—The minimum amount of change for any tree was computed with a modified version of the algorithm of Smith (1992), which was originally designed to compute Euclidean Steiner trees in multiple dimensions (trees that have minimal length in terms of the sum of Euclidean distances along the branches; ”minimum evolution” according to Cavalli-Sforza and Edwards 1967; Thompson 1973). Smith’s algorithm uses a branch-and-bound approach to find the shortest unrooted, bifurcating tree connecting the data points. Instead of Smith’s routines that optimize the positions of internal nodes to minimize the sum of the amounts of shape change on all branches measured as Euclidean distances, our version optimized the internal nodes to minimize the sum of squared distances. As a result of the branch-and-bound strategy, the algorithm is guaranteed to find the topology of the tree that provides the minimal sum of squared changes, and it is quicker than an exhaustive search (Swofford et al. 1996; Felsenstein 2004). The Java
Maximal amount of change.—The maximal amount of change was computed for the star phylogeny, a completely unresolved “bush” with a single internal node. Under squared-change parsimony, this internal node can be reconstructed as the mean of the phenotypic values of the terminal taxa. Accordingly, the maximal amount of shape change was computed as the sum of squared Procrustes distances of the 9 species means from the overall mean shape.

RESULTS

The first 2 principal components accounted for 42.5% and 29.1% of the variation among the species means. Plots of these 2 principal components (Fig. 6) therefore show nearly three-quarters of the total variation among species means in just 2 dimensions. The first principal component features a gradient from the more rounded wings of *D. sechellia* to the narrower and more elongated wings of *D. teissieri*. The second principal component represents more complex changes of the shape of the wing tip, the arrangement of the cross-veins, and the landmarks of the wing base.

The reconstruction of evolutionary changes of shape on the combined phylogeny (Fig. 5a) shows a clear divergence between the 2 main clades, primarily in the direction of the second principal component. This is visible in Figure 6 as a separation of these 2 lineages in the upper and lower halves of the plot. Moreover, within each lineage, there appears to be diversification in the direction of the first principal component (the horizontal direction in Fig. 6). It is difficult to judge, from this graph, whether this is homoplasy by parallel evolution because the projection onto the plane of the plot may give the appearance that branches are parallel even if they are at an angle to each other in the multidimensional space. This visual inspection therefore suggests that there is a clear phylogenetic signal and that there may be some homoplasy.

The comparison with the alternative topologies is mostly consistent with this pattern (Fig. 7). If the tree is rooted so that the clade of *D. orena* and *D. erecta* is the sister group to the remaining species (Fig. 7b,d), most of the shape divergence not only is contained within that bigger clade but also suggests a strong phylogenetic structure in the shape variation.

The permutation test confirmed the impression that there was phylogenetic structure in the data. For most topologies and for weighted as well as unweighted

**FIGURE 6.** Reconstruction of evolutionary changes in wing shape. The phylogenetic tree has been superimposed onto a plot of the first 2 principal components of the covariance matrix among species means. This is the projection of the multivariate shape space onto a plane that best represents the variation among means (71.6% of the total variance among species means). The tips of the terminal branches are at the locations of species means (with bootstrapped 95% confidence ellipses). The positions of the internal nodes were reconstructed by squared-change parsimony using the composite tree topology (Fig. 5a) and branch lengths derived from estimated divergence times. The outline diagrams show the shape change from the reconstructed common ancestor (gray outlines and empty circles) to the mean shape for the respective species (black outlines and solid dots), exaggerated 3-fold for better visibility.
squared-change parsimony, the test found a phylogenetic signal that was statistically significant (Table 1). The distributions of tree lengths for all 362,880 permutations were unimodal and more or less symmetrical and did not appear to differ in any major aspects (not shown). The 2 versions of the tree topology for the

<table>
<thead>
<tr>
<th>Phylogeny</th>
<th>Unweighted</th>
<th>Weighted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tree length</td>
<td>P</td>
</tr>
<tr>
<td>Composite tree</td>
<td>0.00237</td>
<td>0.016</td>
</tr>
<tr>
<td><em>period</em> gene, re-rooted</td>
<td>0.00234</td>
<td>0.012</td>
</tr>
<tr>
<td><em>period</em> gene, original rooting</td>
<td>0.00230</td>
<td>0.009</td>
</tr>
<tr>
<td><em>Amyrel</em> gene, re-rooted</td>
<td>0.00249</td>
<td>0.041</td>
</tr>
<tr>
<td><em>Amyrel</em> gene, original rooting</td>
<td>0.00245</td>
<td>0.032</td>
</tr>
</tbody>
</table>

Note: The tree topologies correspond to those of Figure 5. Rooted trees were used for these analyses.

*Figure 5a; tree assembled from several sources (Lemeunier and Ashburner 1984; Harr et al. 1998; Lachaise et al. 2000; Ting et al. 2000; Cariou et al. 2001; Parsch 2003; Tamura et al. 2004; Drosophila 12 Genomes Consortium 2007).

*Figure 5b; re-rooted version of the tree from Lachaise et al. (2000).

*Figure 5c; tree from Lachaise et al. (2000).

*Figure 5d; re-rooted version of the tree from Cariou et al. (2001).

*Figure 5e; tree from Cariou et al. (2001).
Phylogenetic Signal in Morphometric Data

Our example analysis revealed a clear phylogenetic signal in the landmark data on *Drosophila* wing shape (Fig. 4). The data suggest mostly divergent evolution of wing shape among the 9 species included in this study, but there may also be some parallel trends within the 2 main lineages (Fig. 6). The permutation test indicates that this phylogenetic signal is statistically significant (Table 1). The phylogenetic structure in the data is further underscored by the fairly high values of the consistency and retention indices for shape (Table 2).

The permutation test presented here is an extension of the test used by Laurin (2004) for scalar measurements. A significant result of this permutation test, as it was obtained for our example, is a first indication of phylogenetic structure in the data. The null hypothesis of the test is the total absence of any phylogenetic signal, and rejection of this null hypothesis therefore only implies that there is some degree of structure according to the phylogenetic history. This means that phylogeny needs to be taken into account in the analyses, for instance, by using phylogenetic comparative methods (e.g., Harvey and Pagel 1991; Rohlf 2001), but it does not necessarily imply that there is a particularly tight association between the phylogeny and the shape data. If there is no phylogenetic signal, for instance, if adaptation to current environmental factors overwhelms the phylogenetic effects, the use of comparative methods is unnecessary. An example of this is the sexual shape dimorphism of

**DISCUSSION**

Testing the link between morphometric variation and phylogeny has long been a goal for studies of the evolution of shape. Many investigators have compared phylogenetic trees estimated from other data to trees constructed from morphometric distances among the mean shapes of terminal taxa using methods such as the Fitch–Margoliash algorithm, UPGMA clustering, neighbor-joining, or maximum likelihood (Loy et al. 1996; Marcus et al. 2000; Cannon and Manos 2001; Pretorius and Scholtz 2001; Cole et al. 2002; Cardini 2003; Guill et al. 2003; Houle et al. 2003; Cardini and O’Higgins 2004; Lockwood et al. 2004; Moraes et al. 2004; Caumul and Polly 2005; Couette et al. 2005; Larson 2005; Leinonen et al. 2006; Macholán 2006; Neustupa and Škaloud 2007; Cardini and Elton 2008; Frédérick et al. 2008; Panchetti et al. 2008). A different approach is to use squared-change parsimony or related methods for mapping morphometric data onto phylogenetic trees from other sources (Klingenberg and Eka 1996; Corti et al. 1998; Rohlf 2002; Miller and Venable 2003; Nicola et al. 2003; Linde et al. 2004; Stayton 2005; Macholán 2006; Schlick-Steiner et al. 2006; Stayton and Ruta 2006; McPeek et al. 2008; Scholtz et al. 2008; Sidlauskas 2008; Astúa 2009). In this study, we have emphasized a different way to study the connection between morphometrics and phylogeny, through a permutation test for the presence or absence of a phylogenetic signal and measures of homoplasy for shape corresponding to those that have been used in traditional cladistics.

**FIGURE 8.** The minimal-length tree for wing shape, computed using squared-change parsimony (with a modified version of the branch-and-bound algorithm of Smith 1992). This tree fundamentally differs from the consensus of phylogenetic trees obtained from molecular and cytological evidence (cf. Fig. 6) and is therefore very likely to misrepresent the phylogenetic relationships among these 9 species.
Drosophila wing shape, where the permutation test did not indicate a significant phylogenetic signal (based on the same species and the same landmark information as this study; Gidaszewski et al. 2009).

For questions that go beyond the mere presence or absence of a phylogenetic signal in morphometric data, different methods are required. Here we have proposed 2 measures of homoplasy that are derived from the analogous indices used in conventional cladistics (Kluge and Farris 1969; Archie 1989; Farris 1989). The consistency index measures homoplasy by the increase in tree length of the actual tree over the minimal tree length needed for the shape data. The retention index can be interpreted as a measure of synapomorphy as it compares the relative improvements of the actual tree and of the shortest possible tree over the “worst-case” scenario of the star phylogeny (with no synapomorphy by definition).

In our case study for the D. melanogaster subgroup, the fairly high consistency and retention indices indicate that there is relatively little homoplasy in wing shape (Table 2). This result can be interpreted as an expression of evolutionary inertia and is consistent in this respect with the observation that the wing shapes of different species in the genus Drosophila and related genera are remarkably similar (e.g., Houle et al. 2003; Hansen and Houle 2004). This may be surprising, given that the D. melanogaster subgroup is about 5–12 million years old (Russo et al. 1995; Li et al. 1999; Tamura et al. 2004), which would allow large divergences of wing shape even by random drift alone. Moreover, the wing shape of D. melanogaster is evolutionarily quite malleable. It has evolved to form latitudinal gradients on several continents in relatively short time spans (e.g., Gilchrist et al. 2000), and it readily responds to artificial selection in the laboratory (Weber 1990, 1992; Houle et al. 2003). For instance, Houle et al. (2003) applied artificial selection in opposite directions for a local feature of wing shape during 14 generations and obtained a difference between lines that was 15 times as large as the within-line standard deviation. This difference is somewhat in excess of the greatest difference found between any pair of the 9 species in our data set, arisen over many millions of generations of evolution in nature, namely 11.9 times the within-group variation (Mahalanobis distance, a multivariate equivalent to the univariate difference scaled by the within-group standard deviation). This suggests that stabilizing selection has an important role in the evolution of wing shape (Hansen and Houle 2004). Moreover, these findings also suggest that adaptive evolution of wing shape would have had enough time to obliterate the phylogenetic signal in the data. The phylogenetic structure in the data is therefore somewhat surprising and may reflect a similar structure in the selection regimes of the species and not just the slowness of the evolutionary response to selection. This suggestion would have to be tested in a comparative ecological study.

There are no published data on homoplasy of geometric shape that would allow a direct comparison with the results of our case study. The abundance of reversals and convergence in studies that have mapped morphometric data onto phylogenies (e.g., Klingenberg and Ekau 1996; Rohlf 2002; Nicola et al. 2003; Linde et al. 2004; Schlick-Steiner et al. 2006; Sidlauskas 2008) and the often poor fit of phenetic trees derived from morphometric data to independently estimated phylogenetic trees (e.g., Marcus et al. 2000; Pretorius and Scholtz 2001; Cole et al. 2002; Houle et al. 2003; Couette et al. 2005) suggest that greater degrees of homoplasy may be widespread. The measurement of homoplasy is relevant, for instance, in the study of adaptive diversification. Studies of adaptive evolution have focused particularly on instances of parallel evolution in different lineages (e.g., Langerhans and DeWitt 2004; Schluter et al. 2004) and convergent evolution of shape in independent lineages adapting to similar ecological niches (Harmon et al. 2005; Wroe and Milne 2007). Further studies using this approach are therefore desirable.

In this paper, we illustrated the new methods with an example of landmark data. The approach, however, is also applicable to different types of morphometric data. Squared-change parsimony and related methods also have been successfully applied to sets of distance measurements (Klingenberg and Ekau 1996), semilandmarks (Macholán 2006), outline data (Linde et al. 2004; Scholtz et al. 2008), and data on 3-dimensional surfaces (McPeek et al. 2008, 2009). The permutation test for phylogenetic signal and the shape consistency and retention indices can be used for those data types as well.

Morphometric Data and Inferring Phylogenies

Despite the clear phylogenetic signal in the wing shape variation, all methods failed to infer the consensus phylogeny. The topology of the minimal-length tree obtained from the analysis of wing shape (Fig. 8) and the trees obtained with the Fitch–Margoliash and neighbor-joining algorithms or with maximum likelihood were all very different from the relationships that Laurin (1996) studied 9 sea urchin taxa and found between any pair of the 9 species in our data set, arisen over many millions of generations of evolution in nature, namely 11.9 times the within-group variation (Mahalanobis distance, a multivariate equivalent to the univariate difference scaled by the within-group standard deviation). This suggests that stabilizing selection has an important role in the evolution of wing shape (Hansen and Houle 2004). Moreover, these findings also suggest that adaptive evolution of wing shape would have had enough time to obliterate the phylogenetic signal in the data. The phylogenetic structure in the data is therefore somewhat surprising and may reflect a similar structure in the selection regimes of the species and not just the slowness of the evolutionary response to selection. This suggestion would have to be tested in a comparative ecological study.

There are no published data on homoplasy of geometric shape that would allow a direct comparison with the results of our case study. The abundance of the shape data. The retention index can be interpreted as a measure of synapomorphy as it compares the relative improvements of the actual tree and of the shortest possible tree over the “worst-case” scenario of the star phylogeny (with no synapomorphy by definition).

In our case study for the D. melanogaster subgroup, the fairly high consistency and retention indices indicate that there is relatively little homoplasy in wing shape (Table 2). This result can be interpreted as an expression of evolutionary inertia and is consistent in this respect with the observation that the wing shapes of different species in the genus Drosophila and related genera are remarkably similar (e.g., Houle et al. 2003; Hansen and Houle 2004). This may be surprising, given that the D. melanogaster subgroup is about 5–12 million years old (Russo et al. 1995; Li et al. 1999; Tamura et al. 2004), which would allow large divergences of wing shape even by random drift alone. Moreover, the wing shape of D. melanogaster is evolutionarily quite malleable. It has evolved to form latitudinal gradients on several continents in relatively short time spans (e.g., Gilchrist et al. 2000), and it readily responds to artificial selection in the laboratory (Weber 1990, 1992; Houle et al. 2003). For instance, Houle et al. (2003) applied artificial selection in opposite directions for a local feature of wing shape during 14 generations and obtained a difference between lines that was 15 times as large as the within-line standard deviation. This difference is somewhat in excess of the greatest difference found between any pair of the 9 species in our data set, arisen over many millions of generations of evolution in nature, namely 11.9 times the within-group variation (Mahalanobis distance, a multivariate equivalent to the univariate difference scaled by the within-group standard deviation). This suggests that stabilizing selection has an important role in the evolution of wing shape (Hansen and Houle 2004). Moreover, these findings also suggest that adaptive evolution of wing shape would have had enough time to obliterate the phylogenetic signal in the data. The phylogenetic structure in the data is therefore somewhat surprising and may reflect a similar structure in the selection regimes of the species and not just the slowness of the evolutionary response to selection. This suggestion would have to be tested in a comparative ecological study.

There are no published data on homoplasy of geometric shape that would allow a direct comparison with the results of our case study. The abundance of
congruence between a cladogram estimated from qualitative characters and trees computed from Procrustes distances. Analyses of Procrustes distances for temporal bone shapes of hominoids provided correct estimates of the phylogeny when analyzed with the neighbor-joining and Fitch–Margoliash methods (Lockwood et al. 2004). Similarly, González-José et al. (2008) found that parsimony and likelihood trees from data on skull shape in extant and fossil hominoids indicated that the genus *Homo* was monophyletic, although there was conflict about other relationships. Trees derived from 3 sets of morphometric distances of cranial cartilages of tadpoles by UPGMA clustering were mostly consistent with molecular phylogenetic trees (Larson 2005). MacLeod (2002) conducted an analysis of relative warp scores that was able to recover the correct phylogeny from simulated data that had been generated as independent character changes without homoplasy (Naylor 1996). Numerous other studies, however, have revealed partial or complete incongruence between phylogenetic trees obtained from independent data and trees obtained from geometric morphometric data, regardless of whether they were based on UPGMA clustering (Cournet et al. 1997; Marcus et al. 2000; Pretorius and Scholtz 2001; Cole et al. 2002; Milne and O’Higgins 2002; Cardini 2003; Guill et al. 2003; Cardini and Elton 2008; Panchetti et al. 2008; Young 2008), neighbor-joining (Couette et al. 2005; Macholán 2006; Cardini and Elton 2008), or maximum likelihood (Cannon and Manos 2001; Caumul and Polly 2005; Cardini and Elton 2008). On balance, these direct comparisons suggest that geometric morphometric data may not be reliable indicators of phylogeny. Abundant homoplasy may be one reason why morphometric traits often fail to indicate the correct phylogenetic tree, but there are also some difficulties in principle.

Estimation of phylogeny from quantitative traits is usually based on models such as Brownian motion (explicitly so for maximum likelihood methods; e.g., Felsenstein 2004). These models make a number of assumptions, such as isotropic variation, that are rarely met by morphometric data (Klingenberg and Monteiro 2005). But even if all assumptions are met, there is still the problem that the relationship between divergence times and phenotypic differences is inherently very noisy. Using a neutral model, Lynch (1989) determined that the morphological difference between a pair of species diverging from a common ancestor follows a chi-square distribution with a coefficient of variation of approximately 1.4 (i.e., a distribution that is skewed to the right, where the standard deviation exceeds the mean by about 40%). As a result of this large variability, the inference from the morphological divergence between 2 species to the time of separation of the corresponding evolutionary lineages can only be made with a great deal of uncertainty. Consequently, even if the model of Brownian motion holds, it is quite likely that estimates of phylogeny based on morphological divergence are incorrect. The clear phylogenetic signal that is inherent in the Brownian motion process therefore does not assure that trees can be estimated correctly. A possible solution would be to combine information about the shapes of multiple parts, provided that their shapes are independent of each other (this is likely to be difficult in practice because organismal structures are usually integrated, even if a degree of modularity exists; Klingenberg 2008b). Deviations from the model of Brownian motion, such as effects of selection, may also weaken or eliminate the expected relationship between time of divergence and morphological difference, thus reducing the phylogenetic signal, and therefore may further increase the probability that estimates of phylogenetic trees are incorrect. The results from the empirical studies cited above, where a majority of trees from morphometric data conflict with trees from other sources, are consistent with this view.

There is a widespread view that a strong phylogenetic signal in morphometric data (or other phenotypic data, for that matter) implies that the phylogeny can be reconstructed from the morphometric data (e.g., Cole et al. 2002). This is not necessarily the case, as our example demonstrates. To make this relationship more intuitively clear, it may be helpful to use a simile from daily life. A review of a concert, if it is written by a good music critic, will provide a rich and multifaceted account of the performance and will therefore carry a strong “musical signal.” Nevertheless, nobody would expect to be able to reconstruct the entire experience of the concert from reading the review. Similarly, it should not be expected that morphometric data, even if they carry a strong phylogenetic signal, are sufficient to reconstruct the phylogeny of the clade in question. The reconstruction of phylogenetic trees is not the sole criterion for assessing the phylogenetic signal in morphometric data, and a wider range of approaches therefore should be used for that purpose. The permutation test for the presence or absence of a phylogenetic signal and the shape consistency and retention indices, which we have described in this paper, are 2 additional approaches to assess phylogenetic signals in morphometric data.

**Funding**

Financial support for this research was provided by the Faculty of Life Sciences of The University of Manchester, the Biotechnology and Biological Sciences Research Council, and by the Natural Environment Research Council.

**Acknowledgments**

We thank Michel Baylac for discussions in the planning phase of the project. Michel Baylac, Jean-François Ferveur, Laurence Monti, and the late Daniel Lachaise provided flies used in this study, Marie-Louise Cariou supplied extra information on phylogenies, and Paul O’Higgins gave us access to an unpublished manuscript. We are grateful to Andrea Cardini, David Polly, Norm MacLeod, Rod Page, Jack Sullivan, and the anonymous reviewers for their stimulating
comments and criticism, which helped us to improve the manuscript.

REFERENCES


