Phylogenetic morphometrics (I): the use of landmark data in a phylogenetic framework

Santiago A. Catalano[a,b,*], Pablo A. Goloboff[a,c] and Norberto P. Giannini[a,d]

[a]*Consejo Nacional de Investigaciones Científicas y Técnicas;* [b]*Fundación Miguel Lillo, Miguel Lillo 251, 4000 S.M. de Tucumán, Argentina;* [c]*Facultad de Ciencias Naturales e Instituto Miguel Lillo, Miguel Lillo 205, 4000 S.M. de Tucumán, Argentina;* [d]*Programa de Investigaciones de Diversidad Biológica Argentina, Facultad de Ciencias Naturales, Miguel Lillo 205, 4000 S.M. de Tucumán, Argentina*

Abstract

A method for the direct use of aligned landmark data (2D or 3D coordinates of comparable points) in phylogenetic analysis is described. The approach is based on finding, for each of the landmark points, the ancestral positions that minimize the distance between the ancestor/descendant points along the tree. Doing so amounts to maximizing the degree to which similar positions of the landmarks in different taxa can be accounted for by common ancestry, i.e. parsimony. This method requires no transformation of the aligned data or the results: the data themselves are the x, y, z coordinates of the landmarks, and the output of mapping a character onto a given tree is the x, y, z coordinates for the hypothetical ancestors. In the special case of collinear points, the results are identical to those of optimization of (continuous) additive characters.

© The Willi Hennig Society 2010.

The field of geometric morphometrics comprises a vast array of methods to describe shape quantitatively. Geometric morphometrics has important ramifications for many subdisciplines of biology (morpho-functional studies, comparative anatomy, anthropology, etc.), but very limited interactions with phylogenetics. Many of the methods of geometric morphometrics have been developed by pheneticists, and perhaps (as noted by MacLeod, 2002, p. 101) this is the reason why many systematists tend to equate morphometrics with phenetics, and many morphometricians have very little interest in phylogeny. The purpose of this paper is to show that, contrary to common belief, a parsimony framework can accommodate certain types of geometric morphometric data perfectly.

The most widely used methods in geometric morphometrics are based on landmarks. Landmarks are discrete anatomical loci that can be recognized as "the same" in all specimens under study (Zelditch et al., 2004). A configuration is a group of landmarks chosen to represent the shape of a particular structure. Superimposition methods align two landmark configurations (filtering out differences due to rotation, translation and size). The differences in shape are determined from the changes in the relative position of each individual landmark (Benson et al., 1982; Siegel and Benson, 1982; Rohlf and Slice, 1990; Rohlf and Marcus, 1993; Slice, 1996; Rohlf, 1999; Adams et al., 2004). This approach to comparing shapes is appropriate for pairwise comparisons, but insufficient for evaluating shape changes along a phylogeny—for that, a method that explicitly evaluates change between observed taxa and hypothetical ancestors is needed. This paper proposes such a method, based on the parsimony criterion, as a natural three-dimensional extension of Farris optimization.

Description of the approach

Ancestral positions for a single landmark

The method proposed here works on specimens that are already aligned (see Alignment for a discussion of

*Corresponding author:
E-mail address:* sacatalano@gmail.com

alignment criteria). The position of an individual landmark is specified by values along two or three dimensions (in contrast with the unidimensionality of a continuous character). How to establish ancestral states for a character that changes in three dimensions has never been proposed. The most logical approach is to use a 3D extension of Farris optimization, or spatial optimization, choosing for each ancestral point the positions that minimize the displacement D of this landmark along all ancestor/descendant pairs. Let $d(p, q)$ be the Euclidean distance between p and q, and b the number of branches in the tree.[1] Thus the coordinates for the point in each ancestor have to be chosen such that the sum D of displacements:

$$D = \sum_{n=1}^{b} d\left(n, \text{ancestor}_{(n)}\right)$$

is minimum. Such a spatial optimization is perfectly akin to standard parsimony analysis. For continuous characters, Farris optimization (Farris, 1970; Goloboff et al., 2006) finds the values for the ancestral nodes that minimize ancestor/descendant differences (Fig. 1a). In two or three dimensions instead, where the position of each point in space is defined by coordinates x, y, z, the coordinates for ancestral points that minimize the sum of distances between the ancestor/descendant points must be found (Fig. 1b). When the points are collinear, and thus can be represented by a single continuous character (Fig. 2), the spatial optimization becomes identical to Farris optimization (both numerically and conceptually).

For a node in the tree that connects an ancestor and two descendants (A, B, C, already positioned), the point P (see Fig. 1b) which minimizes the sum of distances to the three vertices of the triangle ABC can be calculated analytically. This point is known as the Fermat point (so called because Fermat posed this problem as a challenge to Torricelli, who solved it geometrically in the early 1600s) or first isogonic center (so-called because, in triangles with all angles below 120°, ∠APB = ∠APC = ∠BPC = 120°). For three points, the "tree" formed by the segments AP, BP and CP is the same as the Euclidean Steiner tree. For more than three points connected to a single internal node (a polytomy), the point P which minimizes the sum of distances is known as the geometric median, and cannot be found by analytical means (requiring heuristics).

Finding the point positions that, for a given tree, minimize the total displacements between all ancestor/descendant points for the given landmark is
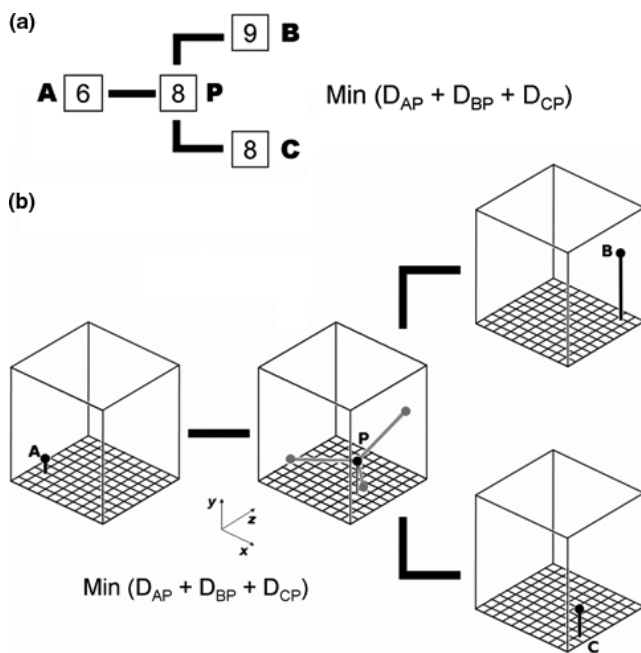

Fig. 1. Comparison between (a) Farris optimization and (b) 3D landmark optimization. In Farris optimization, the sum of the numerical differences between states is minimized. In landmark optimization, the distances between landmark positions are minimized.

[1]the Euclidean distance between two points p, q is $[(p_x - q_x)^2 + (p_y - q_y)^2 + (p_z - q_z)^2]^{1/2}$, which in the case of two dimensions reduces to $[(p_x - q_x)^2 + (p_y - q_y)^2]^{1/2}$.
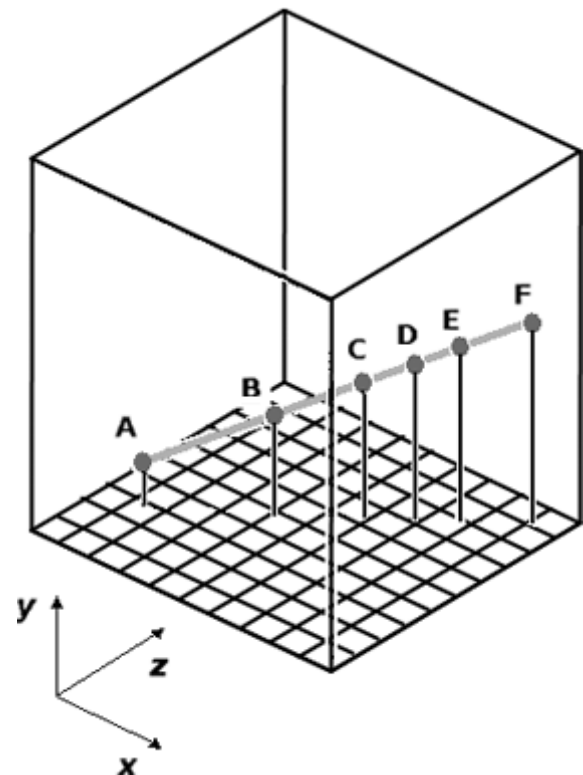

Fig. 2. A set of collinear points. In this case, landmark optimization produces the same results as Farris optimization.

considerably more difficult than for only three nodes, even for binary trees. Suffice it to say here that Goloboff and Catalano (in press) have developed and incorporated into TNT (Goloboff et al., 2003, 2008) heuristic methods (based on a Sankoff approximation as a first step, and a final iterative refinement), which produce a good approximation to the optimum. The problem is related to the problem of Euclidean Steiner trees. The tree that can optimally account for the landmark positions will place all terminals with points in similar positions together, and thus the sum of all ancestor/descendant displacements in the optimal tree for the landmark in question will be the same as the span of the Euclidean Steiner tree for the points of the terminals. The Euclidean Steiner tree thus provides the minimum possible cost for a single landmark.

Ancestral landmark configurations

If the spatial optimization described above is applied to obtain the ancestral position for each of the landmarks in a configuration, then the ancestral "shape" arises as a by-product. Given that the degree of difference between configurations is measured as the sum of linear distances between individual landmarks, obtaining an ancestral configuration by minimizing the displacement of each individual landmark is formally equivalent to calculating ancestral configurations by choosing those that minimize the degree of difference between ancestral–descendant configurations. Note that this rationale used to establish ancestral landmark configurations and quantify the differences among hypothetical and observed landmark configurations is identical to the rationale behind existing superimposition methods (e.g. Siegel and Benson, 1982; Rohlf and Slice, 1990). The partial dependence that may exist between landmarks requires special precautions when using landmarks as data to select a phylogeny (see Tree choice), but has no effect on the shapes resulting from the optimization.

Note that when there is ambiguity in the ancestral positions for several landmarks (i.e. more than one position minimizes change at some node, which is equivalent to ambiguous optimization and multiple most parsimonious reconstructions in the case of discrete characters), some of the combinations of positions may be more defensible reconstructions of ancestral shape, even if equally optimal according to the parsimony criterion employed on a point-by-point basis. Consider as an example several points along a straight line in all observed specimens, a line that can be placed at different heights in different species. If there is ambiguity in the placement of the landmarks in some ancestor, with equally optimal positions at different heights, then placing some points at a high position and others at a lower one will produce a less realistic shape (a zig-zag line) than placing all points at the same height (a straight line). The implementation in TNT considers one "landmark character" to consist of several points, with the idea that the user can specify whether the maximum or minimum x, y or z values should be chosen simultaneously for all points of the "character" in the case of ambiguity. The cases where such a precaution is needed, however, seem to be uncommon in real data sets, as ambiguity arises very infrequently.

Tree choice

For phylogeneticists, the goal of using landmark data will be to help resolve phylogenetic relationships, an issue complicated by the potential dependence between landmarks.

If all the landmarks of a configuration were completely independent, no special treatment would be required: each of the n landmarks should first be optimized independently, with the final score of the tree calculated as the sum of the scores of individual landmarks. If, instead, the dependence was complete, this would amount to giving the configuration a weight n times higher than deserved. Within the realm of discrete characters, this is just like using a character for the colour of each of the P petals of a flower—which amounts to giving the colour of the flower a weight of P, requiring at least P contradicting characters to not have in the most parsimonious tree a group defined by flower colour. But if the colour of each petal is scored with a separate character of weight $1/P$, then the sum of weights for each character representing "flower colour" is effectively the same as that of the single discrete character, and will produce the same result as considering the flower as a whole. This idea can be extended to the case of landmarks, so that the more landmarks are used to describe a shape, the lower the weight each individual landmark should have.

Therefore the steps to perform a phylogenetic analysis from aligned specimens will be (i) optimize each landmark, finding optimal ancestral positions, (ii) calculate its score, (iii) divide the score of each landmark by the number of landmarks in the configuration, (iv) sum the scores of each configuration (possibly taking into account scale and/or units of measurement; see Goloboff and Catalano, in press), (v) sum this score with the remainder of the scores of other characters to obtain the final score of the tree.

The main problem with weighting landmarks inversely to the number of landmarks in the configuration is that the dependence between landmarks in a configuration is evidently not complete, so that down-weighting inversely to the number of landmarks would amount to give the landmark configuration a lower weight than it deserves. Even if perhaps making landmark characters less influential than they should be, this still is a

significant improvement in face of the standard approach of not considering landmark configurations at all, effectively giving them a weight of zero.

An example

An example of the use of spatial optimization to choose from among trees and choose ancestral landmark configurations is shown in Fig. 3. In this artificial example, taken from Naylor (1996) and MacLeod (2002), fish shapes were "evolved" on a "model" tree. The shapes were scanned from MacLeod (2002: p. 127, fig. 7.11) and separated into 12 different configurations following MacLeod (2002). For comparability with previous analyses, the specimens were aligned under Procrustes generalized least squares with tpsRelw (Rohlf, 2008). To decrease the influence on tree choice of those configurations represented by many landmarks, configurations were down-weighted in inverse proportion to the number of landmarks. Landmark rescaling and optimization was performed as described by

Goloboff and Catalano (in press). Searches were done with TNT, using multiple random trees as starting points for tree bisection–reconnection (TBR). Each of the starting points recovered the same optimal tree, identical to the model tree (Fig. 3). In addition, mapping the landmarks onto the optimal tree correctly recovers the changes in landmark positions and thus the synapomorphies—the derived positions of each landmark on a given node.

Alignment

The method described in this paper assumes that the specimens are already aligned. For the alignment, it is advisable to use a method in line with the criterion used to establish ancestral landmark positions—a method that minimizes the sum of the linear distances. Several existing methods for alignment are based on approximate solutions to this problem. Larsen (2008), for instance, proposed a method that approximates an alignment based on Euclidean distances by averaging
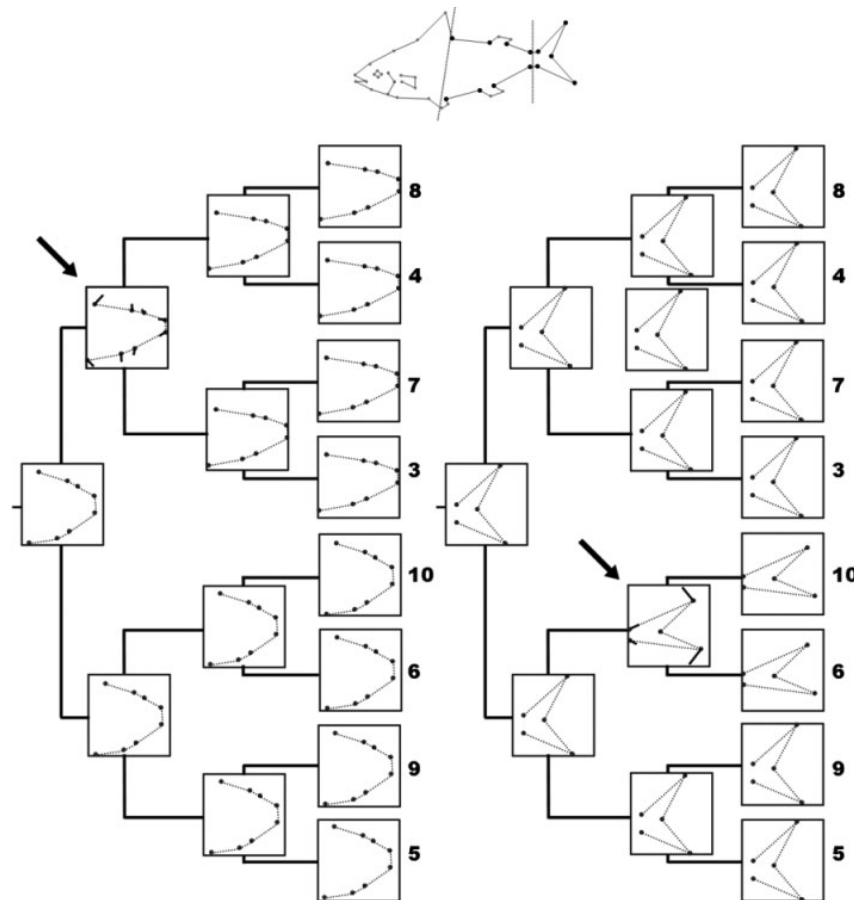

Fig. 3. Ancestral landmark configurations obtained using the spatial optimization for two characters derived from Naylor (1996) and MacLeod (2002). The arrows indicate tree nodes with a landmark displacement relative to their ancestor, i.e. synapomorphies.
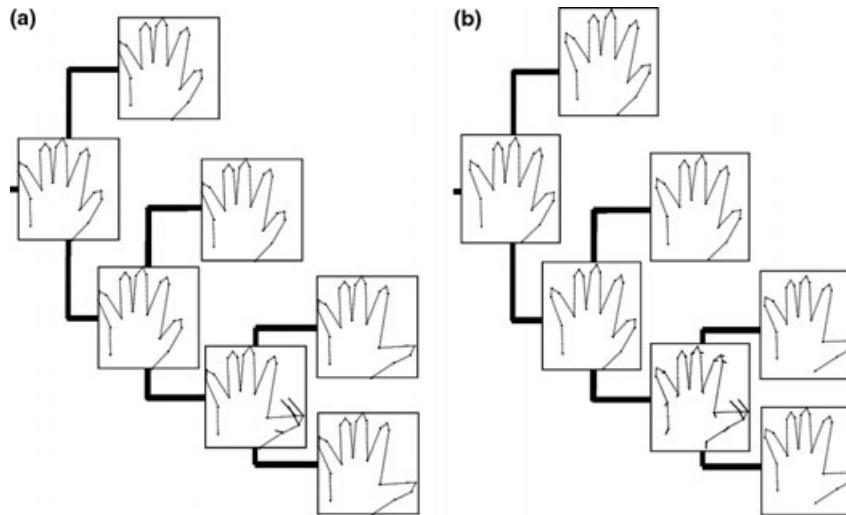
Fig. 4. A hypothetical example showing ancestral landmark positions calculated with spatial optimization, showing the difference between alignments produced with (a) resistant fit theta-rho analysis or (b) generalized least-squares Procrustes alignment. The changes in (b) occur at the expected node, but are spread over all the landmarks instead of only the landmarks representing the thumb.

the Manhattan distance of a series of coordinate systems with equal angular rotation. Other methods, such as the resistant fit theta-rho analysis (RFTRA; Siegel and Benson, 1982; Rohlf and Slice, 1990; Slice, 1996), also produce alignments much more similar to those produced by minimizing the linear distances than those produced by the least-squares criterion. Both methods have additional advantages over least-squares fitting. Least-squares fitting may result in a misleading representation of shape differences (differing more from biological expectation) when one or a few of the landmarks exhibit extreme displacements relative to the others in the configuration (Siegel and Benson, 1982; Rohlf and Slice, 1990; Slice, 1996; Larsen, 2008). Although an alignment based on minimizing linear Euclidean distances will generally be the best method, it is possible that in some circumstances a two-point registration alignment (Bookstein, 1982, 1991) might be preferable.

The example of Fig. 4 (inspired by the example of Larsen, 2008, fig. 5) shows the effect of the alignment in the representation of change along a tree. In Fig. 4a the configurations were superimposed using RFTRA, while in Fig. 4b they were superimposed using least-squares.[2] In both cases, changes onto the tree were inferred by spatial optimization. As expected, the alignment strongly affects the distribution of inferred changes. For the least-squares alignment, the inferred changes were spread over all the landmarks, but for RFTRA all

the changes were inferred to occur in the thumb—the only moving digit. It is already known that the use of least-squares alignment is defensible under some circumstances, for example when all the shapes are expected to be identical, with the differences in landmark position resulting from measurement errors. Since these errors are likely to be evenly spread among different landmarks, it makes sense to use a method of alignment—like least-squares—which will uniformize the displacement among all landmarks. This situation is precisely the opposite of what is expected of landmark data to be used for phylogenetic inference—so it is doubtful that least-squares alignments will ever be desirable for phylogenetic analyses.

As is known (e.g. Benson et al., 1982), superimposition methods in general have limitations in the accuracy with which minor details can be considered. For example, in Fig. 4 the different positions that the landmarks of the thumb can take are restricted to a circular trajectory, but superimposition methods represent change only as rectilinear displacements. Spatial optimization succeeds in recovering the occurrence of the movement, but not its most likely trajectory. Eventually, one might want a modified method such that certain regions of the space are forbidden in some taxa and/or ancestors; that would require additional calculations, and has not yet been implemented. While perhaps realistic, the incorporation of those prohibited subspaces would detract from the simplicity of the method, and at the very least should be used sparingly.

Although the method described here starts from a multiple alignment (in the same way as all methods proposed to date to establish phylogenetic relationships

---

[2]Least-squares alignment was performed using tpsRelw. RFTRA alignment was performed using a C program written by the first author.
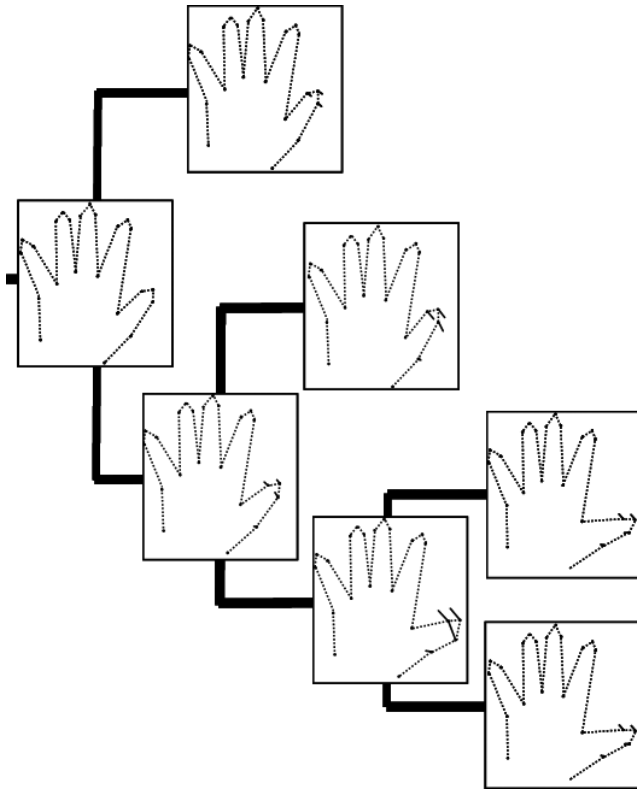
Fig. 5. The same alignment as Fig. 4a, but with ancestral landmark positions calculated using least-squares "parsimony" instead of spatial optimization. The changes are spread over all tree nodes.

from landmark data: Zelditch et al., 1995; MacLeod, 2002; Rohlf, 2002; Lockwood et al., 2004; González-José et al., 2008), such a two-step procedure (first align, then establish phylogenetic relationships) is probably neither unavoidable nor the best way to analyse landmark data. It would, in principle, be possible to analyse this type of data without the need of a multiple alignment, just as in direct optimization of sequence data (Wheeler, 1996). The only difference from the method proposed here is that in a dynamic approach, the configuration of the terminals would also have to be rotated, translated or sized in order to minimize the linear distances (although size changes would have to be constrained, as otherwise the best scores would always be obtained when shapes are shrunk completely). The computational requirements for such a method would be much more important than those needed for fixed alignments and, consequently, the method proposed in this paper can be considered as a heuristic solution to the more general problem.

On the supposed incompatibility between morphometrics and cladistics

The idea that morphometric characters cannot be used in cladistic studies is rather widespread, but this stems

mostly from misconceptions of the logical basis of phylogenetic analysis. The approach presented here is strictly cladistic, in the sense that it evaluates trees based on specific and complete ancestor reconstructions. As in standard parsimony, ancestral points are chosen with the aim of minimizing differences between ancestors and descendants—as in additive (Farris, 1970), non-additive (Fitch, 1971), step-matrix (Sankoff and Rousseau, 1975), or direct optimization (Wheeler, 1996; Varón et al., 2010).

Continuous characters and distances

Several authors (e.g. MacLeod, 2002; Adams et al., 2004; Lockwood et al., 2004) have asserted that cladistics is incapable of dealing with the continuous variation inherent in morphogeometric data, but in fact the specification of an ancestor may be done by assigning a value of a continuous variable or by specification of coordinates for landmark points, exactly as it can be done by specifying a discrete state or a DNA sequence. It is true that phylogeneticists have seldom used continuous characters in practice, but this is due to lack of implementation more than any theoretical impossibility or incompatibility; in fact (as noted by Goloboff et al., 2006), the first algorithm for character optimization (Farris, 1970) was described to work on continuous characters.

One of the proposed solutions (e.g. Lockwood et al., 2004; Couette et al., 2005) for the supposed inability of cladistics to deal with continuous variation is to use pairwise distances as input. This, however, suffers from the same problems as pointed out by Farris (1981, 1985) and Huson and Steel (2004): the tree-scores may well correspond to the distances between abstract, physically unrealizable ancestors. An example (modified from Farris, 1981) is three taxa A, B, C, and a single landmark with coordinates $A = (1,1)$, $B = (2,1)$, $C = (1.5,1.866)$, so that the distance between any two taxa is exactly 1 (A, B and C form an equilateral triangle with each side of length 1). Creating a hypothetical ancestor, H, which is connected to each of the terminals by a branch of length 0.5, produces a perfect fit to the data, but there is no value that the x, y coordinates could take for the landmark in ancestor H such that the displacement from each of the terminals is 0.5. The length of the branches connecting A, B and C through H is then nothing but a mathematical abstraction, with no possible physical realization, and it is hard to see why such a quantity is to be minimized (Farris, 1981).

Homology and ordering

Bookstein (1994) also argues for the basic incompatibility between morphometrics and cladistics on several grounds, the first of which is superficially different from

the arguments about the inability to deal with continuous characters. He claims that homology is a binary, all-or-none definition, because two features either are, or are not, homologous. Bookstein's characterization of cladistics as requiring all-or-none homology assessments is perhaps correct in the sense of all alternative states of a character in a matrix being homologous at some level (since they all must be transformations of some original state),[3] but certainly not at the level of homology of the alternative conditions within a character. The evidence for phylogenetic inference is primary homology (sensu de Pinna, 1991), or similarities, that can both vary by degrees and be ambiguous. Cladists have long accepted the fact that, just as observations of relative degrees of similarity are used to establish homology in the first place, the same type of information can—and should—be used to establish relative degrees of homology within characters (Maslin, 1952; Lipscomb, 1992). In other words, the length of a 5.5-cm-long wing is closer to the length of a 6-cm-long wing than it is to the length of 50-cm-wing—which provides a perfectly logical basis for the decision of considering transformations between 5.5 and 6 cm to be less costly than transformations between 5.5 and 50 cm. It is hard to see how Bookstein may have overlooked the distinction, because aside from being obvious, it is how pheneticists had treated characters—and the way most cladists still do. If Bookstein's characterization were true, then cladistic analysis of continuous (one-dimensional) characters would simply be impossible, but parsimony—as discussed above—works just the same when the alternative conditions for ancestors consist of continuous values instead of discrete states. At heart, Bookstein's claim about the reliance of cladistic analysis on all-or-none homology assessments for the individual states of a character is simply a different version of the same criticism made by Lockwood, MacLeod, Adams and their co-authors: the myth that cladistic analysis cannot deal with continuous characters.

Another argument made by Bookstein is that "no two changes of the 'state' of shape are ever quite the same", which has the implication—for cladistics—that "morphometric shape variables cannot form a hierarchy" (Bookstein, 1994, p. 215). Bookstein's argument is hardly relevant for cladistic analysis, which maximizes explanation of observed similarities, not explanation of changes or transformations (see Farris, 1983, p. 18; Farris, 2008, p. 829; contra Kluge and Grant, 2006). The last of Bookstein's arguments concerns the "shape nonmonotonicity theorem", which asserts that given several shapes, any ordering is arbitrary, because the shapes could always be ordered in many different ways, depending on the reference system or criteria used for

_____
[3]Sequence optimization may not require homology even at that level.

sequencing. Cladistics is the attempt to explain perceived character similarities by postulating inheritance from common ancestors, but Bookstein is discussing whether degrees of similarity themselves can be perceived non-arbitrarily. Thus Bookstein refers not so much to cladistic analysis itself, but to the very observations that a cladistic analysis attempts to explain. Bookstein's argument about the ambiguity in state ordering applies equally well to discrete characters: even treating a character as non-additive or "unordered" amounts to asserting that no two states are closer (and thus to giving all character state transformations equal cost). Were one to consider that Bookstein's "shape nonmonotonicity theorem" is an argument to abandon phylogenetic analysis of morphometric data, one would also have to abandon phylogenetic analysis of discrete characters, as well as most of comparative biology.

Multidimensionality

One of the main concerns about the use of landmark data in cladistic analyses (Rohlf, 1998; Monteiro, 2000; Klingenberg and Monteiro, 2005) is its multidimensionality: the dimensions of the space of all possible landmark configurations, equal to the number of landmarks multiplied by the number of dimensions on which the landmarks are gathered. Each point of this space represents a different landmark configuration. Of course, some configurations with no biologically meaningful differences may differ only in rotation or translation—once those differences have been sorted out, the number of meaningfully different configurations is reduced. It is in the process of excluding those irrelevant differences that the "shape" of this multidimensional space is defined. Restricted or not, these spaces simply refer to all the possible configurations that can be considered as different.

Such a multidimensional space is not exclusive to landmark characters. If several continuous characters are analysed, the space defined is also multidimensional (with dimension equal to the number of characters): each point of this space represents a particular combination of values for each character. Hence, prior to spatial optimization, what precluded a cladistic treatment of landmark data was the multidimensionality of the space where each individual landmark can be positioned (with position determined by x, y, z coordinates), rather than the multidimensionality of the space where landmark configurations are represented.

Spatial optimization differs radically from previous methods, in that it maintains the original individuality of the landmarks throughout the analysis. Other approaches (e.g. those based on partial warps, Rohlf, 2002; or principal component analysis scores, González-José et al., 2008) also start from the unavoidable step of

first considering individual landmarks, but then differ in using the entire configuration of landmarks as a "unit" that is subsequently "projected" along different axes of a multidimensional space, effectively representing all the landmarks in a configuration by a single point in a multidimensional space. What is used for the cladistic analysis is then "characters" representing the values of the projections of this point along different axes of that multidimensional space. The ancestors are obtained by assigning optimal values to those "characters", finally translating back from the axes of projections into the original 3D aligned coordinates. Different methods choose different axes of projections for the configurations, but the basic idea is the same. The main problem with that approach is that minimizing changes along each axis of projection does not have any special meaning in terms of the landmark positions themselves, and consequently the ancestral values obtained for the landmarks do not have any meaning in terms of parsimony.

The Euclidean distance used here was also one of the measures used in phenetic analysis (Sokal, 1961). The crucial difference from phenetic usage is that the Euclidean distance between taxa is a measure representing the "distance" between taxa in the abstract mathematical space defined in the "dimensions" of the different characters. The Euclidean distance of spatial optimization is, instead, the actual distance between the landmark points in the 3D physical space[4] of length, width and height of the aligned shapes. Further, when characters change along one dimension only (e.g. collinear points), the Euclidean distance between points summed across characters is exactly the same as the Manhattan distance, as used by Farris (1970). Failing to appreciate this difference has also been the basis for some arguments (Bookstein, 1994; Adams and Rosenberg, 1998; Rohlf, 1998, 2002; Monteiro, 2000; MacLeod, 2002; Adams et al., 2004) about cladistic methods being inappropriate for analysing morphogeometric data. Rohlf (1998), for example, stated that:

> "Methods of statistical analysis should not give different results dependent upon different choices of the orientation of the reference shape. This means that the usual linear parsimony method (Farris, 1970) should not be used to estimate ancestral states since computations minimizing Manhattan distances are not invariant to the effects of rotation."

The rotation of the reference system or the reference shape determines different *x*, *y*, *z* coordinates for each

---

[4]One of the reviewers stated that our distinction "is mistaken because, following the Procrustes superimposition, landmarks are not in a 3D physical space...the distance between landmarks is not physically meaningful". We disagree, and so do Bookstein (1985) and Rohlf and Slice (1990), who "emphasize that the distance measures used in morphometrics should represent actual physical distances on the objects" (Rohlf and Slice, 1990: 56).

point, but both the spatial optimization as defined here and the method of Farris (1970) are invariant to this rotation, even when they are "Manhattan" in the sense of being a sum across characters. The rotation to which spatial optimization and the linear parsimony method of Farris (1970) are sensitive is, instead, the "rotation" of the multidimensional space represented by different characters, but this hardly constitutes a problem for phylogeneticists: such a space is not "rotated" as part of a cladistic analysis.

Additional considerations

On the meaning of homoplasy

One property of the method presented here is that the branch lengths along the path between two terminal taxa (or patristic distance) may well be greater than the observed distances, even for three-taxon problems. Farris (1967) provided one of the earliest quantitative definitions of homoplasy, as the departure of patristic distances from observed distances. Under that definition, our approach implies that any three points that are not collinear always require some "homoplasy", even for only three taxa. In standard cladistics, the internal node of three-taxon problems always can be assigned a state such that there is no homoplasy. What is necessarily absent from three-taxon problems, however, is not the homoplasy as the difference between patristic and pairwise distances, but instead homoplasy as steps beyond the minimum possible. This is in fact an alternative—and more widely used—meaning of homoplasy. The equivalence between the two definitions of homoplasy does not hold for some types of character. Consider three terminal taxa, each with a different nucleotide (A, C and G) in a non-additive character. The optimal assignment for the internal node will be either A, C or G, any of which implies two steps. However, any of these states implies that there is a pair of taxa for which the patristic distance exceeds the pairwise distance—e.g. choosing A for the internal node, the sum of branch-lengths along the path from the taxon with state C to the taxon with state G equals two, while the observed distance between those two taxa is only one. Does this imply that the tree requires homoplasy for these three taxa? The minimum possible among all trees is two steps, so the assignment of either state to the internal node can be considered to imply no homoplasy (i.e. no steps beyond the minimum possible number of steps with which the character could have evolved), and this is so despite the difference between patristic and observed distances. The two meanings become equivalent only under the type of characters Farris (1967) was considering (binary or additive).

In the case of Sankoff characters, the minimum possible steps may well be dependent on the state(s) to be considered as possible candidates for the interior nodes. Consider the example provided in the documentation of TNT (Goloboff et al., 2003): states A, C, G and T, with all costs defined as 10, except any changes to or from T, defined as 6. The observed terminals have only states A, C and G, and the minimum possible cost in that situation depends on whether or not the (unobserved) T is considered as a candidate for the assignment. If T is a possible candidate, then trees with 18 steps can be obtained, but if T is not a valid candidate, no tree can have fewer than 20 steps (we leave it to the reader to verify that this is the case). In the present approach, instead, any point in the 2D or 3D space will be considered as a valid candidate for any given internal node, so that the shortest tree (for a given landmark) will always have a length equivalent to the span of a Steiner tree.

Squared-changes "parsimony" of x, y coordinates

One of the approaches to establishing ancestral shapes is the independent optimization, with squared-changes parsimony, of the projections along the x, y axes of the landmark. This is implemented in Mesquite (Maddison and Maddison, 2009) and MorphJ (Klingenberg, 2008). One of the properties of the minimization of the squared differences of x, y coordinates is that it also minimizes the squared distances between the landmark points themselves, making the method independent of the system of coordinates used to indicate landmark positions. This independence, however, is not sufficient for a method to be without problems: squared-changes parsimony "perversely ascribes change where none is required and certainly does not minimize ad hoc hypotheses of homoplasy [incorrectly implying] widespread homoplasy on the cladogram" (Hormiga et al., 2000, p. 444)—since it does not maximize the number of similarities explained by common ancestry, "parsimony" is a misnomer for this method. This problem obviously affects the optimization of landmark coordinates as well: if the method is incapable of dealing appropriately with one-dimensional variables, it cannot be expected to provide magically a good solution for higher-dimensional problems. The problem of least-squares minimization for landmark data is illustrated in Fig. 5. While least-squares infers changes all over the tree, spatial optimization (true parsimony) correctly recovers all the changes as synapomorphies of the clade A–B; the only metric that can properly evaluate the degree to which the similarity in landmark position in the terminals can be accounted for by common ancestry is (as in Farris optimization) a linear distance.

The fact that the results of minimizing squared-changes in x, y coordinates are independent of the coordinate system does not mean that the results of independent minimization of the linear differences—true parsimony—along the x, y coordinates will also be independent of the reference system; they are not (as shown in Fig. 6). The only way to minimize displacement of points changing position in a 2D or 3D space is by explicitly considering the actual positions of the points in 2D or 3D space—no reduction to one-dimensional optimization is possible.

Applicability

Landmark studies typically refer to low taxonomic levels. Just like other methods for most parsimonious reconstructions of ancestral states, the present method will always find some reconstruction for any data set. A four-taxon analysis with the shapes of a beetle, a starfish, a cow and a whale will produce some results, but they will obviously not be very meaningful. Landmarks are likely to carry genealogical information, and to be strictly comparable, only for a certain degree of phylogenetic relatedness. Likewise, the existence of a method for cladistic analysis of landmarks does not mean, in any way, that discrete or one-dimensional characters lose their meaning or relevance. Spatial
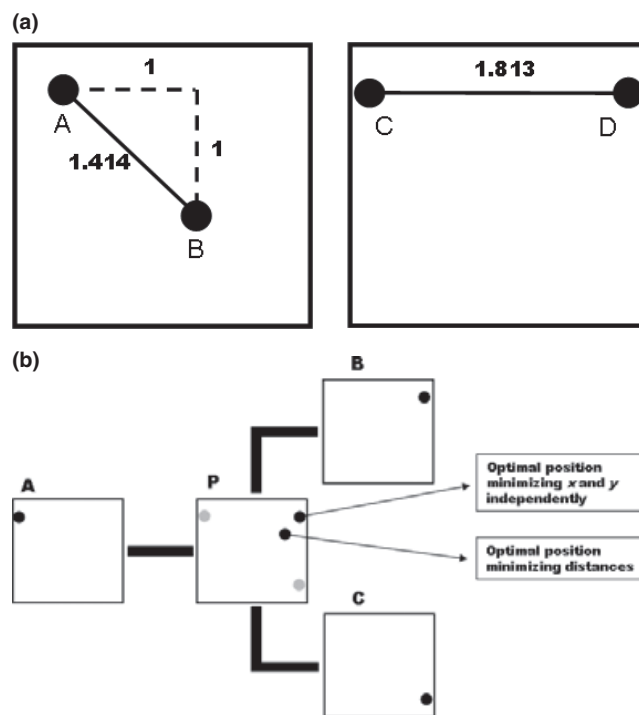

Fig. 6. (a) The costs assigned to a change in position when x, y coordinates are independently optimized are not necessarily ranked according to the spatial distance between the points. In the example, the cost (A–B) is lower than the cost (C–D) when the distances in space are considered, but larger when x, y values are optimized independently. (b) The position obtained by minimizing differences in x, y independently does not minimize point distances.

optimization is in no way a replacement for any of the existing standard procedures in cladistics. As noted by Goloboff et al. (2006, p. 600), if the observed variation can be meaningfully partitioned in discrete states, it is best to avoid more complex analyses like those presented here.

Conclusions

The approach presented here provides a rationale for using morphogeometric data in phylogenetic studies. From the phylogenetic point of view, the approach is entirely equivalent to standard parsimony analysis: it seeks the ancestral landmark configurations that minimize point displacements between ancestral/descendant nodes along all branches of the tree. This is also equivalent to maximizing the degree to which similar positions of the landmark in different taxa can be accounted for by common ancestry, i.e. parsimony. From a morphogeometric point of view, our approach evaluates changes of shape in the same way as all superimposition methods do—changes between shapes are represented by changes in the positions of individual landmarks. Our method can be considered as an extension of those methods to determine changes of shape, but taking into account not only observed shapes, but also ancestral hypothetical shapes, as implied by phylogenetic relationships. Spatial optimization, unlike previous methods that attempted to adapt morphometric data to cladistic studies, requires no transformation of the aligned data or the results: the data themselves are the x, y, z coordinates of the landmarks, and the output of mapping a character onto a given tree is simply (a diagram with) the x, y, z coordinates for the hypothetical ancestors. This simplicity is in keeping with the spirit of parsimony as a logically sound method of phylogenetic reconstruction.

Acknowledgements

References

Adams, D., Rosenberg, M., 1998. Partial warps, phylogeny, and ontogeny: a comment on Fink and Zelditch (1995). Syst. Biol. 47, 168–173.

Adams, D., Rohlf, F., Slice, D., 2004. Geometric morphometrics: ten years of progress following the 'revolution'. Ital. J. Zool. 71, 5–16.

Benson, R.H., Chapman, R.E., Siegel, A.F., 1982. On the measurement of morphology and its change. Paleobiology 8, 328–339.

Bookstein, F.L., 1982. Foundations of morphometrics. Annu. Rev. Ecol. Syst. 13, 451–470.

Bookstein, F., 1985. Transformations of quadrilaterals, tensor fields, and morphogenesis. In: Antonelli, P. (Ed), Mathematical Essays on Growth and the Emergence of Form. University of Alberta Press, Alberta, pp. 221–265.

Bookstein, F., 1991. Morphometric Tools for Landmark Data. Geometry and Biology. Cambridge University Press, New York.

Bookstein, F., 1994. Can biometrical shape be a homologous character? In: Hall, B.K. (Ed.), Homology: The Hierarchical Basis of Comparative Biology. Academic Press, New York, pp. 197–227.

Couette, S., Escarguel, G., Montuire, S., 2005. Constructing, bootstrapping, and comparing morphometric and phylogenetic trees: a case study of New World monkeys (Platyrrhini, Primates). J. Mammal. 86, 773–781.

Farris, J., 1967. The meaning of relationship and taxonomic procedure. Syst. Zool. 16, 44–51.

Farris, J., 1970. Methods for computing Wagner trees. Syst. Zool. 19, 83–92.

Farris, J., 1981. Distance data in phylogenetic analysis. In: Funk, V., Brooks, D. (Eds.), Advances in Cladistics. New York Botanical Garden, Bronx, New York, pp. 3–23.

Farris, J., 1983. The logical basis of phylogenetic analysis. In: Platnick, N.I., Funk, V.A. (Eds.), Advances in Cladistics II. Columbia University Press, New York, pp. 7–36.

Farris, J., 1985. Distance data revisited. Cladistics 1, 67–85.

Farris, J., 2008. Parsimony and explanatory power. Cladistics 24, 825–847.

Fitch, W., 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. 20, 406–416.

Goloboff, P., Catalano, S.A., in press. Phylogenetic morphometrics (II): algorithms for landmark optimization. Cladistics (in press).

Goloboff, P., Farris, J.S., Nixon, K., 2003. TNT: Tree Analysis Using New Technology. Program and documentation available at http://www.zmuc.dk/public/phylogeny.

Goloboff, P., Mattoni, C., Quinteros, A., 2006. Continuous characters analyzed as such. Cladistics 22, 589–601.

Goloboff, P., Farris, J.S., Nixon, K.C., 2008. TNT, a free program for phylogenetic analysis. Cladistics 24, 774–786.

González-José, R., Escapa, I., Neves, W., Cúneo, R., Pucciarelli, N., 2008. Cladistic analysis of continuous modularized traits provides phylogenetic signals in *Homo* evolution. Nature 453, 775–778.

Hormiga, G., Scharff, N., Coddington, J., 2000. The phylogenetic basis of sexual size dimorphism in orb-weaving spiders (Araneae, Orbiculariae). Syst. Biol. 49, 435–462.

Huson, D., Steel, M., 2004. Distances that perfectly mislead. Syst. Biol. 53, 327–332.

Klingenberg, C., 2008. MorphoJ. Faculty of Life Sciences, University of Manchester, UK. http://www.flywings.org.uk/MorphoJ_page.htm.

Klingenberg, C.P., Monteiro, L.R., 2005. Distances and directions in multidimensional shape spaces: implications for morphometric applications. Syst. Biol. 54, 678–688.

Kluge, A., Grant, T., 2006. From conviction to anti-superfluity: old and new justifications of parsimony in phylogenetic inference. Cladistics 22, 276–288.

Larsen, R., 2008. L1 Generalized Procrustes 2D shape alignment. J. Math. Imaging Vis. 31, 189–194.

Lipscomb, D., 1992. Parsimony, homology and the analysis of multistate characters. Cladistics 8, 45–65.

Lockwood, C., Kimbel, W., Lynch, J., 2004. Morphometrics and hominoid phylogeny: support for a chimpanzee–human clade and differentiation among great ape subspecies. Proc. Natl Acad. Sci. USA 101, 4356–4360.

MacLeod, N., 2002. Phylogenetic signal in morphometric data. In: MacLeod, N., Forey, P. (Eds.), Morphology, Shape and Phylogeny. Taylor & Francis, London, pp. 100–138.

Maddison, W., Maddison, D., 2009. Mesquite: a modular system for evolutionary analysis. Version 2.6. http://mesquiteproject.org.

Maslin, T., 1952. Morphological criteria of phyletic relationships. Syst. Zool. 1, 49–70.

Monteiro, L., 2000. Why morphometrics is special: the problem with using partial warps as characters for phylogenetic inference. Syst. Biol. 49, 796–800.

Naylor, G., 1996. Can partial warp scores be used as cladistic characters? In: Marcus, L.F., Corti, M., Loy, A., Naylor, G., Slice, D.E. (Eds.), Advances in Morphometrics. Plenum, New York, pp. 519–530.

de Pinna, M., 1991. Concepts and tests of homology in the cladistic paradigm. Cladistics 7, 367–394.

Rohlf, F., 1998. On application of geometric morphometrics to studies of ontogeny and phylogeny. Syst. Biol. 47, 147–158.

Rohlf, F.J., 1999. Shape statistics: procrustes superimpositions and tangent spaces. J. Class. 16, 197–223.

Rohlf, F., 2002. Geometric morphometrics and phylogeny. In: MacLeod, N., Forey, P. (Eds.), Morphology, Shape and Phylogeny, Systematic Association Special Volume Series 64. Taylor & Francis, London, p. 4.

Rohlf, F., 2008. tpsRelw. Version 1.46. Department of Ecology and Evolution, State University of New York, Stony Brook, New York.

Rohlf, F., Marcus, L.F., 1993. A revolution in morphometrics. Trends Ecol. Evol. 8, 129–132.

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

Sankoff, D., Rousseau, P., 1975. Locating the vertices of a Steiner tree in an arbitrary space. Math. Program. 9, 240–246.

Siegel, A., Benson, R., 1982. A robust comparison of biological shapes. Biometrics 38, 341–350.

Slice, D.E., 1996. Three-dimensional, generalized resistant fitting and the comparison of least-squares and resistant-fit residuals. In: Marcus, L.F., Corti, M., Loy, A., Naylor, G., Slice, D. (Eds.), Advances in Morphometrics. Plenum Press, New York, pp. 179–199.

Sokal, R., 1961. Distance as a measure of taxonomic similarity. Syst. Zool. 10, 70–79.

Varón, A., Vinh, L., Wheeler, W., 2010. POY version 4: phylogenetic analysis using dynamic homologies. Cladistics 26, 72–85.

Wheeler, W., 1996. Optimization alignment: the end of multiple sequence alignment in phylogenetics. Cladistics 12, 1–9.

Zelditch, M., Fink, W., Swiderski, D., 1995. Morphometrics, homology, and phylogenetics: quantified characters as synapomorphies. Syst. Biol. 44, 179–189.

Zelditch, M., Swiderski, D., Sheets, H., Fink, W., 2004. Geometric Morphometrics for Biologists: A Primer. Elsevier Academic Press, London.