In Phenotypic Integration, 2004 (M. Pigliucci and K. Preston, eds.). pg. 325-344.

PHYLOGENETIC COMPARATIVE ANALYSIS OF MULTIVARIATE DATA

Scott J. Steppan

Department of Biological Science

Florida State University

Tallahassee, Florida 32306-1100

E-mail: steppan@bio.fsu.edu

The Problem: Comparative Analysis of Multivariate Data

A central question in fields ranging from quantitative genetics to phenotypic integration and developmental biology is how do patterns of covariation among traits evolve? The fields of inquiry differ in which underlying causes for the covariation they investigate, but they share the challenge of comparing multivariate data sets. In order to understand the evolution of those patterns, as opposed to their maintenance within an organism, it is critical to do so in a phylogenetic context. That is, history must be incorporated into the analysis. A phylogenetic perspective is only recently gaining much application in the pursuit to understand how organisms are built, how their structures are integrated through genetic organization and developmental processes, and how these integrating features both shape and are shaped by evolution. This chapter will discuss the methods available for comparing patterns of integration among taxa and how to conduct those comparisons phylogenetically.

The comparative analysis of multivariate data presents special challenges. In addition to analyzing and visualizing patterns of variation among individuals within a group, one must simultaneously integrate phylogenetic information. Not only must one identify the patterns present within a group (species or population), one must extract the subtle differences among the patterns. These subtle differences will be distributed among groups (taxa) and will create a pattern associated with the phylogenetic history of the taxa. These patterns within patterns must be analyzed and understood simultaneously with the within-group patterns and thus add another dimension to the data that cannot be identified using conventional statistical analyses (Fig. 1). In addition, phylogenetic comparisons implicitly involve multiple taxa, because just two taxa provide no phylogenetic information. The pattern of differences among taxa thus becomes hierarchical in nature (i.e., reflecting the phylogenetic hierarchy) which extends the inherently hierarchical nature of multivariate data because the developmental program is also hierarchical. These inherent complexities of multivariate data produce the greatest challenges for comparative analyses.

Two sets of issues that must be addressed to analyze multivariate data in a comparative context are how to compare data across taxa and how to incorporate phylogenetic information

into those comparisons. In this chapter I will argue that the phylogenetic approach is critical to identifying shared multivariate features and reconstructing their evolution. I will then summarize some of the special problems with the comparative analysis of multivariate data, survey the most common set of methods used to compare multivariate data, and highlight approaches that incorporate phylogenetic information. Because most studies of phenotypic integration and related questions summarize multivariate data in matrix form, I will focus on comparisons between matrices. Additional discussions of these issues and methods can be found in Chernoff and Magwene (1999) and Cheverud et al. (1989).

This chapter looks at multivariate data, whether intended for study of phenotypic integration or comparative quantitative genetics (Steppan et al. 2002), from a phylogenetic perspective. The focus is understanding how patterns of covariation among traits evolve. This objective requires an independently derived phylogeny, typically estimated from the explosively growing body of DNA sequence data. I will not discuss the use of morphometric or multivariate data to estimate phylogenies. That is an entirely separate issue with its own set of challenges and controversies (Rohlf 1998; Zelditch et al. 1998).

Why phylogeny is important

Some form of phylogenetic information is absolutely necessary if we are to reconstruct the evolutionary history of traits. At its simplest, the information could be derived from a pairwise comparison between taxa, but only if the pair consists of ancestor and descendant, wherein the phylogenetic information establishes a temporal polarity. Any other single-pair comparison merely documents a difference, not how that difference evolved. When conducting multitaxon comparisons, additional information is needed to polarize pairwise comparisons or partition the data into a set of phylogenetically independent comparisons (e.g., phylogenetic independent contrasts; Felsenstein 1985). Those partitions are the clades or ancestor-descendent pairs in a phylogeny. The statistical and philosophical justification for the use of phylogenetic information in comparative studies (hereafter referred to as the comparative method) has been discussed in detail elsewhere (Felsenstein 1985; Brooks and McLennan 1991; Harvey and Pagel 1991; Garland and Adolph 1994). Briefly, multitaxon comparisons without a phylogeny inflate

the degrees of freedom in statistical tests by treating the values of traits measured for each taxon as independent observations when in fact they are not. Taxa may be similar for a given trait not because they evolved that trait independently but because they inherited it from an ancestor. Statistically significant correlations between traits (e.g., an environmental factor and a putative adaptation or two features hypothesized to share developmental origins) can be due entirely to common ancestry rather than common adaptive or genetic mechanisms (Felsenstein 1985). Multitaxon comparisons without a phylogeny are also unable to determine the temporal sequence of events (Coddington 1988), the appropriate taxon comparisons, or which condition is ancestral and which is derived and thus cannot determine either the magnitude or the direction of evolution.

Phylogenetic comparisons of DNA sequences use patterns of variation across nucleotides to estimate bias (transformation rates among nucleotides or variation in such rates among nucleotide sites) and identify gene regions under functional constraint or selectionphylogenetic analysis of phenotypes can uncover the evolutionary lability of traits and of phenotypically integrated features. The remainder of this chapter explores numerous questions that can be addressed effectively only by adopting a phylogenetic perspective within a basic comparative approach, but here I will mention some of the key questions. How evolutionarily labile are covariance traits? (covariance traits are the elements of phenotypic integration or covariance structure, identifiable as the covariances or sets of covariances among traits). What is the rate of change in covariance traits? And, although it does not necessarily require a phylogeny, which traits are labile and which are conserved?.

How multivariate data differ from univariate data in comparative studies

Multivariate data are more than just a set of univariate traits; they include the correlations or covariances among the traits. Mathematically, this is to the difference between a vector of means and a matrix of variances and covariances. Students of phenotypic integration are interested primarily in the patterns and magnitudes of covariances, not simply the means of traits, but the comparative methods that have been developed to date are designed for univariate data, as will be described below. The most common methods test for an association between evolutionary changes of two traits, specifically in their means or modes. The association tested occurs across

time and clades, not among individuals of a taxon. In the phylogenetic comparative method, each trait is typically optimized independently onto a phylogeny (i.e., "mapped"), and then a pairwise comparison or regression is made between the optimizations of two traits. Most optimization involves reconstructing ancestral states so as to minimize some objective function (e.g., number of steps, sum of squares, likelihood) over the entire phylogeny. Optimization is relatively well understood for univariate data (Swofford et al. 1996; Schluter et al. 1997) provided there is no significant directional bias (Cunningham et al. 1998) or if that bias is known a priori. With multivariate data, the variances and covariances of many traits must be simultaneously optimized, and theory is lacking on what model to use for these covariances or how best to do that.

A unique challenge presented by multivariate data is how to simultaneously optimize or reconstruct ancestral matrices when the traits forming them are correlated. Should each correlation and covariance be optimized independently? In addition, the correlations/covariances are themselves correlated (hence the integration that is being studied) and they are correlated to varying degrees. This problem just exacerbates that of accurate ancestral reconstruction (Schluter et al. 1997; Cunningham et al. 1998) , where even standard reconstructions of ancestral character means can be prone to a high degree of error. Continuous traits often have such large optimization errors that confidence intervals for the oldest ancestors can exceed the range of variation observed among all extanttaxa (Schluter et al. 1997).

Multivariate studies place greater demands on sample sizes than do univariate comparative studies because group parameter estimates (variances and covariances) are reduced to single observations, and because the analyses must be able to detect subtle variations among patterns. The matrices must be well estimated from large sample sizes or the analyses may just be interpreting estimation error as signal. The greater sampling requirement may be met for some taxa, but comparative studies often include taxa that are uncommon. This problem should be taken into account when designing research projects.

None of the current techniques for the comparative method address fully the concerns discussed here and most do not address them at all. In the next section I will survey available techniques for matrix comparisons and then in the following section I will discuss applying them

in a phylogenetic framework. Although some modifications and future directions will be noted, for the time being we are faced with incomplete analytical solutions to the problem.

MATRIX COMPARISON TECHNIQUES

The array of analytical techniques for comparing matrices can be bewildering. The choice of technique will depend primarily on the question asked and on the nature and structure of the data. For example, element-by-element comparisons are most appropriate when specific correlations or covariances are of interest, but it is excessively cumbersome in searches for patterns of integration. I will describe several classes of techniques below, highlighting how they can be applied to multiple taxa. Example references and the strengths and weakness of these techniques are briefly summarized elsewhere (Steppan et al. 2002).

Element-by-Element

The element-by-element approach is the most detailed, as it involves taking each element of a matrix in turn and testing for significant differences. When many elements are being tested, adjustments like the Bonferroni correction (Rice 1989) must be made to the significance level before interpretation. Statistical tests can be as simple as a *t*-test in conjunction with resampling techniques (Brodie 1993; Roff 2000), but for complex situations other tests may be preferred, such as Shaw's restricted maximum likelihood (Shaw 1987; 1991). The element-by-element approach does not take into account the nonindependence of matrix elements (covariances); several may be correlated because of changes in an underlying factor (e.g., pleiotropy). For an extreme example, if two matrices are proportional (as when variance/integration has increased population-wide), all corresponding elements in the two matrices will be significantly different from each other but as a result of only one underlying biological cause. This approach is typically used for taxon pairs, but to extend it to multiple taxa, one could ask whether an element varies significantly across many taxa.

Matrix correlation

Matrix correlations are a measure of similarity in pattern that is calculated by testing for correlation between the values of corresponding elements. In effect, the elements are the observations, and the taxa are the variables. Again, elements are considered independent of each other, an assumption that is frequently violated. Proportionality cannot be distinguished from matrix equality, and many different changes to a matrix can result in the same correlation. The latter is a greater problem for matrix correlation than most other techniques described here because it reduces all differences between matrices to a single dimension, whereas some other techniques are able to discriminate between different kinds of changes. For some questions, these shortcomings may be unimportant, as when a general measure of similarity is needed. Used in this way, matrix correlation becomes a distance (similarity) metric, but if the patterns that produce changes are of interest, then other methods are preferable (Steppan 1997a). Matrix permutation methods like Mantel's test (Dietz 1983) can test for significant matrix correlation between taxa. The null hypothesis is no correlation, an uncommon condition for morphological traits, such that rejection of the null can be trivial and uninteresting (see Pigliucci's chapter in this book). If the question of interest is whether significant divergence in patterns of covariation (integration) has occurred, then techniques that assume equality as the null (common principal components analysis, maximum likelihood) are preferable. Multiple taxa cannot be compared directly, but comparisons can be structured hierarchically (see Single Pairs section below).

Maximum Likelihood

Maximum likelihood methods (Shaw 1987; Shaw 1991) can be used to test a wide variety of hypotheses about variance-component matrices. Hypotheses of equality can tested for individual elements, subsets of matrices (see Pleiades below) and entire matrices. With a well-estimated data set, this method permits statistically precise statements about which parts of the matrix differ and by how much. Although not part of the current implementation of maximum likelihood, the great potential of this method might be in its integration with likelihood estimates for ancestral reconstruction (see below) and perhaps even phylogeny reconstruction (Swofford et al. 1996). Models of the evolution of covariance structure could be tested within a likelihood

framework that integrates the likelihood of a given model over the range of possible error in the estimation of covariance elements, uncertainty in ancestral reconstruction, and perhaps even phylogenetic error (Huelsenbeck et al. 2000). This integrative approach with likelihood is computationally demanding but may be within reach in the near future, perhaps by applying a Bayesian statistical framework.

Ordination

Rather than comparing the many elements in a matrix individually, ordination techniques can be used to transform the data according to some optimality criterion and comparisons can be made among these new variables. An objective is often to reduce the dimensionality of the data so that analysis and interpretation are simplified. Ordination can also be effective at data exploration, and principal components analysis, (PCA) is often used for that purpose. Simplification may be the primary objective. The choice of ordination technique should depend on the hypotheses tested. In one example, Voss and colleagues used PCA to identify a size vector so that ontogenies could be compared across populations (Voss et al. 1990), whereas later factor analysis (FA) was used to examine differences in patterns of integration (Voss and Marcus 1992). In PCA or FA, analyses are conducted on each operational taxonomic unit (OTU; whether each OTU is a population of species) separately and then pairwise or other comparisons made between OTU factor loadings. A modification of FA, confirmatory factor analysis (CFA), can be used to compare observed factor structure to a hypothesized factor structure (Zelditch 1987; Zelditch et al. 1990). Although none of these techniques is designed for multiple OTUs, paired comparisons can be structured in groups to test questions in an evolutionary context. An underappreciated concern when applying ordination methods is that the structure in the data being maximized may not match the patterns produced by the biological mechanisms or predicted by the conceptual models being tested. Major causal mechanisms may not result in unrelated, orthogonal vectors of variation (PCA; Houle et al. 2002) or maximized covariance (FA). For example, changes in allometry will likely rotate the orientation of the first principal component in PCA and that may force a rotation in the other eigenvectors even if the localized underlying developmental programs have not changed.

The most notable ordination technique that can incorporate multiple taxa simultaneously is common principal components analysis (CPCA; Flury 1987a; 1987b). In CPCA, a set of matrices is tested for the presence of a common vector or vectors under a PCA model. The number of matrices that can be considered is unlimited. Each level of shared structure is compared to an adjacent level, creating a hierarchical set of tests. The method discriminates among matrices with a wide variety of levels of shared structure, including equality (not significantly different), proportionality (unequal, but the hypothesis of proportional eigenvalues is not rejected), CPC, partial CPC (some but not all eigenvectors in common), and unrelated (no shared structure). There are N-2 levels, where N equals the number of traits. The method has been reviewed repeatedly (Steppan 1997a; Phillips and Arnold 1999; Roff 2000; Steppan et al. 2002) and will not be presented in detail here. The key advantages of CPCA for comparative phenotypic integration studies are its ability to decompose matrix structure into a hierarchical set of structures and its ability to analyze many taxa together under the same model. It can also be applied to developmental data in a variety of contexts (Klingenberg and Zimmermann 1992; Klingenberg et al. 1996). Despite its growing popularity, CPCA also has several deficiencies that should be considered, including the underlying PCA assumption of orthogonality of factors (Houle et al. 2002; Steppan et al. 2002). Additionally, we have a poor understanding of how pleiotropy affects covariance structure, and the detailed and seemingly clear-cut results of CPCA may entice a researcher to be overconfident in the results.

An alternative approach would be to perform ordinations across taxa, where the observations are the matrix elements (variances and covariances) rather than the individual measurements. In this approach, matrices are converted into vectors, creating taxon-by-covariance-element matrices (each column is a taxon, each row is a matrix element) that could themselves be subjected to ordination analyses. Taxa (OTUs) would be grouped by similarity in covariance structure and resultant axes would be determined by the variation in patterns of covariance structure. This method would have three advantages. First, it would allow multitaxon comparisons. Second, examination of the those covariances that are highly correlated with the new factors would identify those suites of covariances that vary most among taxa or help identify

correlation pleiades (see below; Berg 1960). By subjecting them to canonical variates analysis or similar techniques which maximize correlations among sets of both independent and dependent variables, one could identify which suites of covariances are maximally correlated with ecological or behavioral variables and thus test models of adaptive integration. Third, the OTUs could be ordinated or plotted by their similarity in covariance structure ("covariance space"), and OTUs joined graphically according to their phylogenetic relationships. The phylogeny would thus be mapped onto that ordination, in contrast to mapping covariance structure onto the phylogeny, as suggested below.

The taxon-by-covariance-element approach just described is similar in its restructuring of the data to a MANOVA approach advocated by Roff (2002). In Roff's MANOVA method, developed specifically for the difficult statistical problem of comparing genetic variance/covariance matrices, matrix elements become the columns and jackknife pseudoreplicates become the rows (each row is a jackknife run where a single family has been excluded from the analysis). The new matrix follows a MANOVA structure in adding columns that are coded to specify OTU (all OTUs are combined in a single data set) and any environmental variable of interest. Various experimental designs can be used including nested analyses. The relative contributions to the variances/covariances by the variables of interest (e.g., species, phylogeny, sex, environment) can then be partitioned and the significance level of the effect estimated.

Integration Indices

Measures of overall integration, rather than the patterns of integration within matrices, can also be employed (Chernoff and Magwene 1999). Examples include the average absolute values of the correlations (Cane 1993) and the variance of the correlation matrix's eigenvalues (Wagner 1984; Cheverud et al. 1989). These are more difficult perhaps to interpret in a phylogenetic context but can be mapped onto a phylogeny as a single trait. This approach can determine whether some clades display more integration then others, but these methods cannot detect homoplasy in integration. That is, higher integration values can be achieved by different

mechanisms (e.g., increased canalization of different trait complexes), and thus high or low values of integration may not be homologous. Phylogenetic methods assume homology.

Distance

Appropriate distance metrics have been little explored. A careful selection of metrics could be used to explore different aspects of integration or covariance structure that may be shared among taxa. The most common metric is disparity (Willis et al. 1991; Steppan 1997b), the summed off-diagonal, element-by-element differences in a covariance matrix. Variants of disparity have been used as well. Podolsky et al. (1997) used a standardized disparity, GD^2, in which the differences between covariance elements are squared before summing and then the sum is divided by the number of elements times the mean covariance. This latter step was done to standardize the metric when sets of matrices with different numbers of traits are compared, but it would be unnecessary with matrices of the same form. Roff and colleagues included the diagonal elements (variances) and called this version the T method when it was used with permutation to test for significant disparity (Roff et al. 1999; Begin and Roff 2001). Permutation adds an important statistical test to a general approach usually lacking statistics. Compared to disparity, the T method will be more influenced by changes in population variance than covariance structure, because the absolute values of variances and their differences is often much greater than the covariances.

Matrix correlations could be converted to distances by subtracting them from one, but these distance measures would obviously still have some of the limitations mentioned above. Other metrics include correlation disparities (i.e., disparity scaled by variances) or other variations (largely unexplored) chosen to emphasize different aspects of matrix structure. The "size" and "shape" of the differences between matrices could be calculated in a manner analogous to removing size in multivariate studies by regressing corresponding matrix elements, and then calculating disparity among the residuals (i.e., disparity corrected for proportional changes in matrices). It might then be informative to subject the derived distances to further analysis, including ordination analyses like multidimensional scaling (MDS, a dimension-reducing method that estimates new dimensions that minimize the distances between OTUs).

"Pleiades" Approaches

Berg (1960) introduced the concept of correlation pleiades to describe suites of highly correlated traits that are relatively independent of other such suites within an organism. Despite the intuitive appeal of this concept, actually identifying pleiades within a taxon can be difficult since there are few clear criteria for grouping traits either a priori (on the basis of first principles) or a posteriori (based on observed trait correlations). Nevertheless, several different empirical approaches have been developed. Olson and Miller (1958) identified ρ-groups as traits bonded by phenotypic correlations greater than some set value or values. They compared the ρ-groups to *F*-groups, sets of traits hypothesized to be functionally integrated a priori. Cheverud (1982) extended this framework to comparisons with *G*-sets, analogous to ρ-groups but based on genetic correlations. The maximum-likelihood method of Shaw has been extended to determine whether taxa share subsets of a matrix, where subsets are analogous to pleiades. This restricted maximum-likelihood model (REML) has been applied mostly to quantitative genetic parameters (Shaw 1987; Shaw 1991; Service 2000). Unlike CPCA, it has not yet been extended to multiple taxa, but it could be quite powerful. Other approaches are discussed in other chapters of this book and elsewhere (Cheverud et al. 1989; Chernoff and Magwene 1999; Magwene 2001). These include common space analysis (Flury 1987b; Flury 1988), and dimensionality (Kirkpatrick et al. 1990). See Steppan et al. (2002) for brief review).

PHYLOGENETIC APPROACHES TO MULTIVARIATE DATA

Once the patterns of phenotypic integration have been characterized for a set of OTUs, the focal question becomes how to compare these patterns within a phylogenetic context. Extrapolating studies of phenotypic integration beyond single OTUs to phylogenetic comparisons adds three demands. First, multiple comparisons are required, and second, to be phylogenetic, either those comparisons must be structured according to the phylogeny or special phylogenetic analyses must be conducted. As stated before, the latter are noticeably lacking for multivariate data, so much of this discussion will examine means of adding phylogenetic structure. Third, because the objective is to extract subtle differences from the within-taxon covariance structure,

precise estimation of the individual matrices is even more important than is the case for noncomparative studies. I will not discuss this third demand here but raise it as an underappreciated consequence of comparative studies.

Comparative methods can be subdivided into those for discrete data and those for continuous data (Harvey and Pagel 1991). Discrete data, such as presence/absence (Brooks and McLennan 1991), can be analyzed by character mapping (optimizing the character state transformations onto a phylogeny under a given optimality criterion) or sister group comparisons. In the latter, sets of sister groups, where one member of each pair possesses a key trait and the other lacks it, are surveyed for conformity to a hypothesized pattern. The number of observations is equal to the number of sister group pairs. There is a wider variety of methods for continuous data and this type of data is generally of more interest in the field of integration. Continuous methods include minimum evolution, phylogenetic autocorrelation, and independent contrasts and its variants. I will discuss the application of these methods in more detail below.

I have organized the various approaches into five conceptual groups that differ by the manner in which the data are integrated with phylogenetic information: character mapping, single pair, hierarchical group, ancestral reconstruction, and phylogenetic correlation. Graphical examples of three of these approaches are shown in Fig. 2.

Character Mapping

Character mapping is not a multivariate technique but one where discrete or continuous characters are extracted from a matrix and then mapped (optimized) individually onto a phylogeny. An example is shown in Fig. 1, where one can recognize a change in ρ-group membership coinciding with the divergence of the species pair on the left. One would reconstruct the loss of a strong correlation and the gain of another on the branch leading to this clade. If one had hypothesized such a change on the basis of a shift in function of a structure made up by those traits, then the data would support the functional hypothesis, sensu Coddington (1988). Repeated examples from other clades or subclades would strengthen the conclusion. In the example in Fig. 1, the change in group membership represents a single observation, and many more clades must be sampled for statistical power. Of the various ways of characterizing integration, element-by-

element and pleiades approaches are most amenable to extraction of single characters for mapping, but gain or loss of factors from FA could also be mapped (B. Chernoff, pers. comm.). One could also treat covariances or correlations as individual continuous characters and map them, but estimation error could be significant, and minor changes in correlations could be overinterpreted when mapped onto a phylogeny. The following approaches are explicitly multivariate.

Single Pairs

The single pairs approach simply makes all possible pairwise comparisons between taxon matrices (Fig. 2a). Statistically this is the least justifiable approach because it involves repeated comparisons (each of N OTUs is involved in N - 1 comparisons), but it can provide a useful overview of patterns or range of variation for a metric (Steppan 1997a; Ackermann and Cheverud 2000; Marroig and Cheverud 2001). This approach has been used when others are difficult to apply because of the nature of the question being asked or the matrix comparison method. For example, in leaf-eared mice *Phyllotis*, all possible matrix correlations were obtained, then grouped by the taxonomic level at which each was calculated (Steppan 1997a). The comparison of interest was the matrix correlation, a simple measure of overall similarity, but which cannot be used with the hierarchical group analysis (a multigroup method) or autocorrelation and is less appropriate than character-based or distance measures when used with ancestral reconstruction methods (matrix correlation applied to ancestor-descendent pairs would yield similarity values between them; distance measures would be more appropriate). Other analytical approaches such as CPCA can be applied in single-pair fashion when the questions justify it or when the number of taxa is very low (2 or 3), but hierarchical group and ancestral reconstruction methods would be more powerful for larger data sets.

In the single pairs approach, phylogenetic structure is provided by grouping comparisons among taxa according to their taxonomic or phylogenetic level. In Fig. 2a, taxa A, B, and C are in clade (e.g., genus) 1, and D, E, and F are in clade (e.g., genus) 2. A simple categorization would group all comparisons between species of the same genus (shown by the solid arrows) together and all comparisons between different genera (dashed arrows) together. One could subdivide the

within-genus category to distinguish between -sister species comparisons (A-B and E-F) from between -sister clade comparisons (equivalent to second-level sister groups; C-A, C-B, D-E, and D-F). In my study of *Phyllotis* (Steppan 1997a) I referred to a population-level phylogeny for the species group, and the deeper-level comparisons were the progressively more cumbersome levels of sister species, second-sister clades (sister species once removed), and third-sister clades.

An alternative grouping criterion is by formal taxon level, such as species, subgenus, genus, and so on (Marroig and Cheverud 2001). This criterion has the advantages of grouping more comparisons within fewer categorical levels, and it does not require an explicit phylogenetic hypothesis. For most groups, fully resolved phylogenies do not exist, and one must rely on the less resolved systematic classifications. One disadvantage of using formal taxonomy is that not all sister-species pairs in a phylogeny are equally old or divergent. For example, if one sampled one clade (say a genus) much better than its sister group, the sister-species pairs within the better-sampled genus are likely to be closely related, but if the poorly sampled sister clade includes only two species, say from different subgenera or genera, then that species pair could be equivalent in divergence time to third- or fourth-sister species groups in the other genus. Equivalent levels of divergence would not then be grouped for analysis. One solution to this problem is to usethe branching information (topology) from the phylogeny and categorize the level of grouping by formal taxonomy. That was the approach used in Steppan (1997a) and Ackermann and Cheverud (2000). A more sophisticated approach is to incorporate branch-length information, as discussed below. Branch lengths are distinct from branching sequence and comparative biologists have been much slower to incorporate branch length information because only with a very good fossil record or the recent accumulation of molecular data can we have good estimates of them.

A variant comparison strategy was used by Podolsky et al. (1997) wherein only a subset of all possible comparisons was made. For each higher-level comparison, only the most similar pairs between members of sister clades (or UPGMA clusters) were compared. They used this approach to reduce computation time for maximum likelihood analyses, rather than to avoid statistical problems, and it is not the preferred method.

Hierarchical Group Analysis

Hierarchical group analysis may be the most appropriate method for analyses that allow multiple groups but where ancestral state reconstruction is problematic. At present, CPCA (common principal component analysis) is becoming the most popular technique for data with these properties and will be used as the example. In hierarchical group analysis, a CPC analysis is conducted on all members of a clade for each clade in turn (Fig. 2b). Thus, A and B would be analyzed together, as would be E and F. Then A, B, and C would be analyzed together as would D, E, and F. Finally, all taxa would be analyzed at once (Steppan 1997a; Ackermann and Cheverud 2000; Badyaev and Hill 2000). One goal is to identify the branches in the phylogeny at which specific components are no longer shared by all members of a clade. Another goal is the inverse, to identify clades where the hypothesis of common structure can be accepted.

Several questions remain unanswered about the behavior of CPCA with respect to phylogenetic analyses. How does it respond to increasing numbers of taxa? Resampling suggests that rejection of shared structure becomes more likely with more taxa, perhaps because of single outlier matrices (Steppan 1997a). How is CPCA affected by outlier or divergent matrices in an analysis (Steppan et al. 2002)? A further problem is that of repeated comparisons, although the nature of the test (acceptance or rejection of homogeneity of covariance structure) makes this statistical violation only slightly problematic.

Although hierarchical group analysis is more appropriate statistically than single-pair comparisons because it makes fewer repeated comparisons, the latter does provide a complementary perspective. For example in *Phyllotis*, CPCA indicated no shared structure at any level in the PCA hierarchy except between closely related populations, suggesting major divergence in covariance structure; but single-pair matrix correlations found the matrices were still very similar ($r^2 > 0.90$; Steppan 1997a). Ackermann and Cheverud (2000) and Pigliucci and Kolodynska (2002) found a similar situation in *Saguinus* tamarin monkeys and *Arabidopsis*, respectively. With large sample sizes, CPCA may be very powerful at detecting small deviations in covariance structure and thus reject hypotheses of shared structure, while single-pair

comparisons can estimate the magnitude of differences (i.e., significance level is not a measure of magnitude).

Ancestral Reconstruction

The term ancestral reconstruction groups together a variety of comparative methods that use reconstructed ancestral values. The most common types of methods are phylogenetic independent contrasts (PIC; Felsenstein 1985), minimum evolution (ME; Huey and Bennett 1987), and more recently phylogenetic generalized least squares (PGLS; Rohlf 2001). Of these, only minimum evolution has been applied to multivariate data (Steppan 1997b) to my knowledge, although Rohlf (2001) suggests the multivariate extension of these methods that reconstructs variances and covariances could be computationally straightforward. Provided that ancestors can be measured or estimated accurately, this approach is the most valid statistically and uses all of the available phylogenetic information. Comparisons are made either between sister species, be they two extant taxa or ancestors (PIC), or between ancestor descendant pairs (ME). These methods were specifically developed to estimate evolutionary correlations (coevolution) between some environmental factor or proximate phenotypic trait (the independent variable) and an adaptive response to it (the dependent variable; e.g., Huey and Bennett 1987). Their strength is that they partition the comparisons such that all are phylogenetically independent of each other. Additionally, the ME method, essentially a variant of character mapping, can be used to partition multivariate evolution to specific branches on the phylogeny.The ME method polarizes ancestor-descendant comparisons while PIC and PGLS do not.

Nearly all the matrix-comparison methods can be used with the ancestral reconstruction approach. I used the distance metric disparity to examine rates of covariance evolution in *Phyllotis* (Steppan 1997b). When branch lengths were scaled by matrix disparity, I found that branches leading to populations and subspecies clades were significantly longer than deeper branches leading to species and more inclusive clades. This pattern contrasted sharply with those reconstructed from Euclidean distance in means or parsimony estimates of DNA sequence evolution (Steppan 1997b). This result suggests that the tempo of covariance evolution or the shape of the adaptive landscape is qualitatively different for covariance structure than for the

gross phenotype or selectively neutral DNA substitutions (most substitutions were synonymous). Further extensions of this approach could be employed, such as testing for constant rates by the relative rates test or likelihood ratio tests under a "covariance clock," analogous to a molecular clock, using maximum likelihood. It should be noted that the sample sizes for variable characters may be much smaller for multivariate data (i.e., number of covariance elements) than for molecular data (nucleotide sites), and thus rate tests are likely to lack power.

Phylogenetic Correlation

Three methods that use different correlational strategies are grouped under phylogenetic correlation. Cheverud and Dow (1985) and Cheverud et al. (1985) introduced phylogenetic autocorrelation (PA) as a phylogenetic extension of spatial autocorrelation methods. The rationale behind all autocorrelation methods (phylogenetic, spatial, serial) is that observations in close proximity will be relatively similar because they share many processes or histories that are not of direct interest to the investigator. Autocorrelation (actually autoregression) removes the effect of proximity from the overall pattern and identifies the scale over which the underlying (but unknown) processes operate. After removal of proximity effects, the remaining pattern is the focus of the researcher, most often adaptation in comparative studies. The special function of PA is to estimate the proportion of variance across taxa in a trait that is explained by phylogenetic relatedness. The phylogenetic component has been called phylogenetic inertia. PA uses a connection matrix that summarizes phylogenetic relationships among taxa. The connection matrix can be determined from a variety of phylogenetic descriptors, ranging from integers representing taxonomic category (e.g., species, genus, family), to number of nodes separating taxa on a tree, to methods that incorporate branch-length information and the height of the common ancestors above the root (Rohlf 2001). Residuals from the autoregression procedure can be used as new characters after the effects of phylogeny have been removed. Again, modifications to the implementation would be needed to use PA with multivariate data.

An alternative to PA, the phylogenetic eigenvector regression (PVR), regresses trait values against the principal coordinate eigenvectors of a pairwise phylogenetic distance matrix

(Diniz-Filho et al. 1998). The performance of this method has not been widely explored nor has its application to multivariate data.

In one of the few applications of phylogenetic correlation to multivariate data, Marroig and Cheverud (2001) calculated all pairwise matrix correlations for a group of New World monkeys and then built a taxon-by-taxon covariance similarity matrix from them. This covariance similarity matrix was compared with genetic and ecological (dietary) distance matrices to determine that diet was more important than phylogeny in explaining similarity in covariance structure.

Branch Length Information (Rate of Evolution)

Incorporating information on branch lengths allows analyses that could not be conducted without a phylogenetic approach. Fig. 3 illustrates the potential importance of branch-length information. The two phylograms presented have identical topologies but very different branch lengths that are drawn proportional to time. A common assumption in studies of quantitative trait evolution is that the expected variance in a trait increases proportionally to time since divergence (Brownian motion model). Therefore, on the left tree, A and B are very closely related (i.e., have diverged relatively recently), and we would expect them to have very similar covariance matrices. In contrast, on the right tree, A and B shared very little unique history, and we would not expect them to be much more similar to each other than either is to C. Strong trait divergence between A and B would be remarkable on the left tree but unremarkable on the right. Likewise, if A and B were very similar on the right tree but the other four taxa were divergent, this observation would be contrary to expectations and would suggest some mechanism for this conservation, possibly canalization or stabilizing selection. The exception to these expectations would be the case in which one assumes a punctuated model of evolution, in which all change occurs at speciation and no subsequent anagenetic evolution takes place along lineages, making branch lengths irrelevant.

Branch-length information either can be added to or is already an important part of all the phylogenetic approaches discussed above. For example, when the single-pairs approach is used, matrix similarity values can be plotted against or regressed on a continuous character describing relatedness (e.g., height of each pair's common ancestor above the root) rather than their

taxonomic category (e.g., within genus). PICs are often calculated with contrasts scaled by the square-root of the branch lengths assuming a Brownian motion model of evolution.

As outlined above, in addition to the use of independently derived branch lengths (usually from DNA sequence data) to inform comparative multivariate analyses, branch lengths can be calculated from the multivariate data themselves. "Covariance lengths" represent the amount of evolution in covariance structure and when scaled by an independent estimate of branch lengths, they represent evolutionary rates. High rates of covariance evolution appeared associated with population-level processes (sampling error, drift, or morphological divergence of populations/phylogenetic species) rather than with the evolution of putative reproductive barriers between biological species or divergence of clades in *Phyllotis* (Steppan 1997b). This result suggests that variation in covariation structure is becoming "saturated," perhaps because stabilizing selection on developmental pathways limits the divergence of covariance structures much beyond some fluctuation produced by drift. It also suggests the possibility of exploring the mechanisms underlying covariance evolution by means of saturation curves analogous to those plotted for nucleotide sequence data. In saturation curves, covariance distance (e.g., disparity) would be plotted against genetic distance or time. Deviations from a linear (for corrected DNA sequence distances) or quadratic (for time, the expectation under Brownian motion) relationship would give evidence for mechanisms limiting variation. Lineage-specific changes in relative branch lengths (in which all members of a clade share particularly long or short branches as detected by a relative-rates test) could be evidence for accelerated or retarded rates of the evolution of covariance structure. Single anomalously long branches (relative to time) might suggest a reorganization of the developmental or genetic program.

WHAT DO DO?

The comparative analysis of multivariate data presents the researcher with a wide variety of less than optimal choices for methods of analysis. There are two sets of choices to be made. First is how to compare matrices and how to measure differences in patterns of covariation. The choice of analytical method will depend in part on the aspect of covariance evolution of interest,

and each method has strengths and weaknesses. There does not appear to be a clearly superior method for all circumstances, but maximum likelihood, model-based ordination techniques, and distance methods show the most promise for future improvement. Second is how to incorporate phylogenetic information. None of the comparative methods developed explicitly for phylogenetic data are optimized for multivariate data. Again, the choice of phylogenetic approach will vary with the question of interest and all may have their place in a study. Ancestral reconstruction and hierarchical group approaches are perhaps the most powerful and flexible for use with multivariate data. With the surge of molecular phylogenies in the literature, it will become increasingly valuable to include branch length information into the analyses to address both increasingly sophisticated questions and old questions with greater precision.

Acknowledgments

I thank Anne Thistle, Katherine Preston, and Massimo Pigliucci for comments improving the quality of the manuscript. Financial support was provided NSF grant DEB-0108422.

Literature Cited

Ackermann, R. R., and J. M. Cheverud. 2000. Phenotypic covariance structure in tamarins (genus *Saguinus*): a comparison of variation patterns using matrix correlation and common principal component analysis. Am. J. Phys. Anthropol. 111:489-501.

Badyaev, A. V., and G. E. Hill. 2000. The evolution of sexual dimorphism in the house finch. I. Population divergence in morphological covariance structure. Evolution 54:1784-1794.

Begin, M., and D. A. Roff. 2001. An analysis of G matrix variation in two closely related cricket species, Gryllus firmus and G-pennsylvanicus. J. Evol. Biol. 14:1-13.

Berg, R. L. 1960. The ecological significance of correlation pleiades. Evolution 14:171-180.

BRODIE, E. D., III. 1993. HOMOGENEITY OF THE GENETIC VARIANCE-COVARIANCE MATRIX FOR ANTIPREDATOR TRAITS IN TWO NATURAL POPULATIONS OF THE GARTER SNAKE *THAMNOPHIS ORDINOIDES*. EVOLUTION 47:844-854.

BROOKS, D. R., AND D. A. MCLENNAN. 1991. PHYLOGENY, ECOLOGY, AND BEHAVIOR. UNIVERSITY OF CHICAGO PRESS, CHICAGO.

CANE, W. P. 1993. THE ONTOGENY OF POST-CRANIAL INTEGRATION IN THE COMMON TERN, *STERNA HIRUNDO*. EVOLUTION 47:1138-1151.

CHERNOFF, B., AND P. M. MAGWENE. 1999. AFTERWORD, PP. 319-348. *IN* E. C. OLSON AND R. L. MILLER (EDS.), MORPHOLOGICAL INTEGRATION. UNIVERSITY OF CHICAGO PRESS, CHICAGO, IL.

CHEVERUD, J. M. 1982. PHENOTYPIC, GENETIC, AND ENVIRONMENTAL MORPHOLOGICAL INTEGRATION IN THE CRANIUM. EVOLUTION 36:499-516.

CHEVERUD, J. M., AND M. M. DOW. 1985. AN AUTOCORRELATION ANALYSIS OF GENETIC VARIATION DUE TO LINEAL FISSION IN SOCIAL GROUPS OF *RHESUS* MACAQUES. AM. J. PHYS. ANTHROPOL. 67:113-121.

CHEVERUD, J. M., M. M. DOW, AND W. LEUTENEGGER. 1985. THE QUANTITATIVE ASSESSMENT OF PHYLOGENETIC CONSTRAINTS IN COMPARATIVE ANALYSES: SEXUAL DIMORPHISM IN BODY WEIGHT AMONG PRIMATES. EVOLUTION 39:1335-1351.

CHEVERUD, J. M., G. P. WAGNER, AND M. M. DOW. 1989. METHODS FOR THE COMPARATIVE ANALYSIS OF VARIATION PATTERNS. SYST. ZOOL. 38:201-213.

CODDINGTON, J. A. 1988. CLADISTIC TESTS OF ADAPTATIONAL HYPOTHESES. CLADISTICS 4:3-22.

CUNNINGHAM, C. W., K. E. OMLAND, AND T. H. OAKLEY. 1998. RECONSTRUCTING ANCESTRAL CHARACTER STATES: A CRITICAL REAPPRAISAL. TRENDS ECOL. EVOL. 13:361-366.

DIETZ, E. J. 1983. PERMUTATION TESTS FOR ASSOCIATION BETWEEN TWO DISTANCE MATRICES. SYSTEMATIC ZOOLOGY 32:21-26.

DINIZ-FILHO, J. A. F., C. E. R. DE SANT'ANA, AND L. M. BINI. 1998. AN EIGENVECTOR METHOD FOR ESTIMATING PHYLOGENETIC INERTIA. EVOLUTION 52:1247-1262.

FELSENSTEIN, J. 1985. PHYLOGENIES AND THE COMPARATIVE METHOD. AM. NAT. 125:1-15.

FLURY, B. 1988. COMMON PRINCIPAL COMPONENTS AND RELATED MULTIVARIATE MODELS. JOHN WILEY AND SONS, NEW YORK.

FLURY, B. K. 1987A. A HIERARCHY OF RELATIONSHIPS BETWEEN COVARIANCE MATRICES, PP. 31-43. *IN* A. K. GUPTA (EDS.), ADVANCES IN MULTIVARIATE STATISTICAL ANALYSIS. REIDEL, DORDRECHT, THE NETHERLANDS.

FLURY, B. K. 1987B. TWO GENERALIZATIONS OF THE COMMON PRINCIPAL COMPONENT MODEL. BIOMETRIKA 74:59-69.

GARLAND, T., AND S. C. ADOLPH. 1994. WHY NOT TO DO TWO-SPECIES COMPARATIVE-STUDIES: LIMITATIONS ON INFERRING ADAPTATION. PHYSIOL. ZOOL. 67:797-828.

HARVEY, P. H., AND M. D. PAGEL. 1991. THE COMPARATIVE METHOD IN EVOLUTIONARY BIOLOGY. OXFORD SERIES IN ECOLOGY AND EVOLUTION. OXFORD UNIV. PRESS, NEW YORK.

HOULE, D., J. MEZEY, AND P. PALPERN. 2002. INTERPRETATION OF THE RESULTS OF PARTIAL PRINCIPAL COMPONENTS ANALYSIS. EVOLUTION 56:433-440.

HUELSENBECK, J. P., B. RANNALA, AND J. P. MASLY. 2000. ACCOMMODATING PHYLOGENETIC UNCERTAINTY IN EVOLUTIONARY STUDIES. SCIENCE 288:2349-2350.

HUEY, R. B., AND A. F. BENNETT. 1987. PHYLOGENETIC STUDIES OF CO-ADAPTATION: PREFERRED TEMPERATURES VERSUS OPTIMAL PERFORMANCE TEMPERATURES IN LIZARDS. EVOLUTION 41:1098-1115.

KIRKPATRICK, M., D. LOFSVOLD, AND M. BULMER. 1990. ANALYSIS OF THE INHERITANCE, SELECTION AND EVOLUTION OF GROWTH TRAJECTORIES. GENETICS 124:979-993.

KLINGENBERG, C. P., B. E. NEUENSCHWANDER, AND B. D. FLURY. 1996. ONTOGENY AND INDIVIDUAL VARIATION: ANALYSIS OF PATTERNED COVARIANCE MATRICES WITH COMMON PRINCIPAL COMPONENTS. SYST. BIOL. 45:135-150.

KLINGENBERG, C. P., AND M. ZIMMERMANN. 1992. STATIC, ONTOGENETIC, AND EVOLUTIONARY ALLOMETRY: A MULTIVARIATE COMPARISON IN NINE SPECIES OF WATER STRIDERS. THE AMERICAN NATURALIST 140:601-620.

MAGWENE, P. M. 2001. NEW TOOLS FOR STUDYING INTEGRATION AND MODULARITY. EVOLUTION 55:1734-1745.

MARROIG, G., AND J. M. CHEVERUD. 2001. A COMPARISON OF PHENOTYPIC VARIATION AND COVARIATION PATTERNS AND THE ROLE OF PHYLOGENY. ECOLOGY, AND ONTOGENY DURING CRANIAL EVOLUTION OF NEW WORLD MONKEYS. EVOLUTION 55:2576-2600.

OLSON, E. C., AND R. L. MILLER. 1958. MORPHOLOGICAL INTEGRATION. UNIVERSITY OF CHICAGO PRESS, CHICAGO.

PHILLIPS, P. C., AND S. J. ARNOLD. 1999. HIERARCHICAL COMPARISON OF GENETIC VARIANCE-COVARIANCE MATRICES. I. USING THE FLURY HIERARCHY. EVOLUTION 53:1506-1515.

PIGLIUCCI, M., AND A. KOLODYNSKA. 2002. PHENOTYPIC PLASTICITY AND INTEGRATION IN RESPONSE TO FLOODED CONDITIONS IN NATURAL ACCESSIONS OF ARABIDOPSIS THALIANA (L.) HEYNH (BRASSICACEAE). ANN. BOT. 90:199-207.

PODOLSKY, R. H., R. G. SHAW, AND F. H. SHAW. 1997. POPULATION STRUCTURE OF MORPHOLOGICAL TRAITS IN *CLARKIA DUDLEYANA*. II. CONSTANCY OF WITHIN-POPULATION GENETIC VARIANCE. EVOLUTION 51:1785-1796.

RICE, W. R. 1989. ANALYZING TABLES OF STATISTICAL TESTS. EVOLUTION 43:223-225.

ROFF, D. 2000. THE EVOLUTION OF THE G MATRIX: SELECTION OR DRIFT? HEREDITY 84:135-142.

ROFF, D. A. 2002. COMPARING **G** MATRICES: A MANOVA APPROACH. EVOLUTION 56:1286-1291.

ROFF, D. A., T. A. MOUSSEAU, AND D. J. HOWARD. 1999. VARIATION IN GENETIC ARCHITECTURE OF CALLING SONG AMONG POPULATIONS OF *ALLONEMOBIUS SOCIUS*, *A. FASCIATUS*, AND A HYBRID POPULATION: DRIFT OR SELECTION? EVOLUTION 53:216-224.

ROHLF, F. J. 1998. ON APPLICATIONS OF GEOMETRIC MORPHOMETRICS TO STUDIES OF ONTOGENY AND PHYLOGENY. SYST. BIOL. 47:147-158.

ROHLF, F. J. 2001. COMPARATIVE METHODS FOR THE ANALYSIS OF CONTINUOUS VARIABLES: GEOMETRIC INTERPRETATIONS. EVOLUTION 55:2143-2160.

SCHLUTER, D., T. PRICE, A. O. MOOERS, AND D. LUDWIG. 1997. LIKELIHOOD OF ANCESTOR STATES IN ADAPTIVE RADIATION. EVOLUTION 51:1699-1711.

SERVICE, P. M. 2000. THE GENETIC STRUCTURE OF FEMALE LIFE HISTORY IN *D. MELANOGASTER*: COMPARISONS AMONG POPULATIONS. GENET. RES. 75:153-166.

SHAW, R. G. 1987. MAXIMUM-LIKELIHOOD APPROACHES APPLIED TO QUANTITATIVE GENETICS OF NATURAL POPULATIONS. EVOLUTION 41:812-826.

SHAW, R. G. 1991. THE COMPARISON OF QUANTITATIVE GENETIC PARAMETERS BETWEEN POPULATIONS. EVOLUTION 45:143-151.

STEPPAN, S. J. 1997A. PHYLOGENETIC ANALYSIS OF PHENOTYPIC COVARIANCE STRUCTURE. I. CONTRASTING RESULTS FROM MATRIX CORRELATION AND COMMON PRINCIPAL COMPONENT ANALYSES. EVOLUTION 51:571-586.

STEPPAN, S. J. 1997B. PHYLOGENETIC ANALYSIS OF PHENOTYPIC COVARIANCE STRUCTURE. II. RECONSTRUCTING MATRIX EVOLUTION. EVOLUTION 51:587-594.

STEPPAN, S. J., P. C. PHILLIPS, AND D. HOULE. 2002. COMPARATIVE QUANTITATIVE GENETICS: EVOLUTION OF THE **G** MATRIX. TRENDS ECOL. EVOL. 17:320-327.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, AND D. M. HILLIS. 1996. PHYLOGENETIC INFERENCE, PP. *IN* D. M. HILLIS, C. MORITZ AND B. K. MABLE (EDS.), MOLECULAR SYSTEMATICS. SINAUER ASSOCIATES, SUNDERLAND, MASS.

VOSS, R. S., AND L. F. MARCUS. 1992. MORPHOLOGICAL EVOLUTION IN MUROID RODENTS II. CRANIOMETRIC FACTOR DIVERGENCE IN SEVEN NEOTROPICAL GENERA, WITH EXPERIMENTAL RESULTS FROM *ZYGODONTOMYS*. EVOLUTION 46:1918-1934.

VOSS, R. S., L. F. MARCUS, AND P. P. ESCALANTE. 1990. MORPHOLOGICAL EVOLUTION IN MUROID RODENTS I. CONSERVATIVE PATTERNS OF CRANIOMETRIC COVARIANCE AND THEIR ONTOGENETIC BASIS IN THE NEOTROPICAL GENUS *ZYGODONTOMYS*. EVOLUTION 44:1568-1587.

WAGNER, G. P. 1984. ON THE EIGENVALUE DISTRIBUTION OF GENETIC AND PHENOTYPIC DISPERSION MATRICES: EVIDENCE FOR A NONRANDOM ORGANIZATION OF QUANTITATIVE CHARACTER VARIATION. J. MATH. BIOL. 21:77-96.

WILLIS, J. H., J. A. COYNE, AND M. KIRKPATRICK. 1991. CAN ONE PREDICT THE EVOLUTION OF QUANTITATIVE CHARACTERS WITHOUT GENETICS? EVOLUTION 45:441-444.

ZELDITCH, M. 1987. EVALUATING DEVELOPMENTAL MODELS OF INTEGRATION IN THE LABORATORY RAT USING CONFIRMATORY FACTOR ANALYSIS. SYST. ZOOL. 36:368-380.

ZELDITCH, M. L., W. L. FINK, D. L. SWIDERSKI, AND B. L. LUNDRIGAN. 1998. ON APPLICATIONS OF GEOMETRIC MORPHOMETRICS TO STUDIES OF ONTOGENY AND PHYLOGENY: A REPLY TO ROHLF. SYST. BIOL. 47:159-167.

ZELDITCH, M. L., D. O. STRANEY, D. L. SWIDERSKI, AND A. C. CARMICHAEL. 1990. VARIATION IN DEVELOPMENTAL CONSTRAINTS IN *SIGMODON*. EVOLUTION 44:1738-1747.

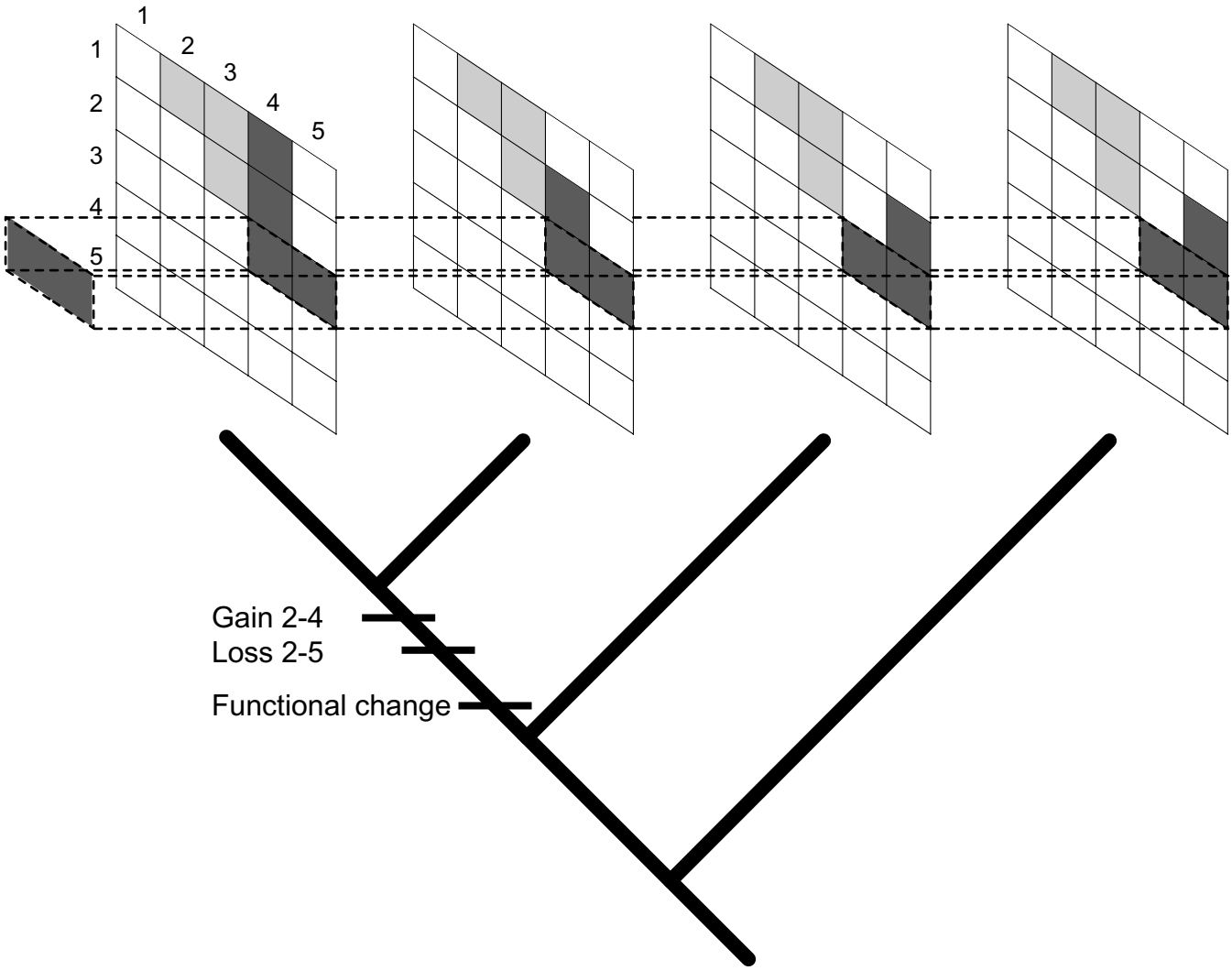
FIGURE LEGENDS

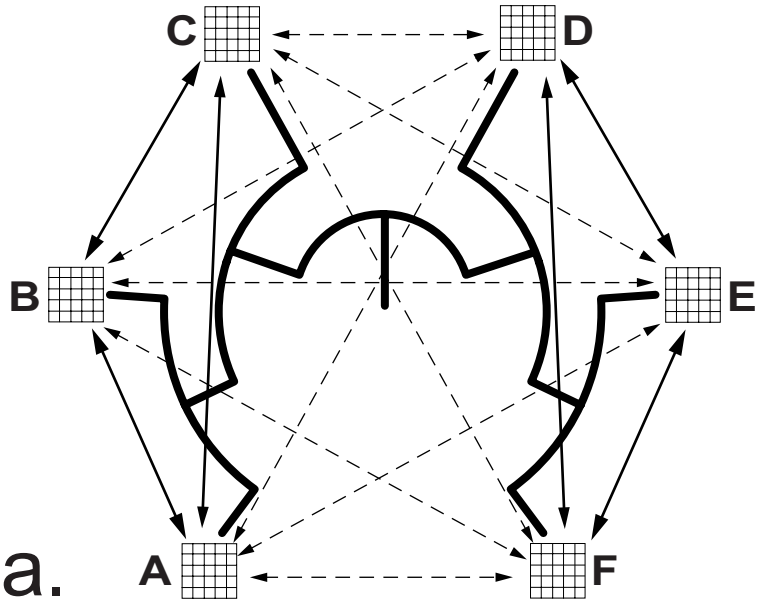
Fig. 1. Hypothetical phylogeny for four species with each species represented by a correlation or variance/covariance matrix for five quantitative traits (rectangles). There are two sets of correlations (covariances) with particularly high values that could be considered correlation pleiades (sensu Berg 1960). These two sets are indicated by the shaded elements. Only sets above the diagonal are shown. Assuming the sets are homologous across all four species, mapping correlations as characters onto the phylogeny results in the conclusion that there have been changes in the membership in the dark set. On the branch leading to the left-hand clade, there has been a loss of the correlation 2-with-5 (2-5) and the gain of the correlation 2-4. Subsequently on the branch leading to the leftmost species, the correlation 1-4 was added to the dark set. Across all four species there is a conserved pair of correlations (3-4, 3-5) that is indicated by the prism transcending the four matrices. See text for discussion of the functional change and its interpretation in a phylogenetic context. The figure illustrates some of the principles involved in identifying both shared patterns of integration and modifications across the evolution of a diversifying lineage.

Fig. 2.Three different approaches to the phylogenetic structuring of comparisons among matrices are illustrated. The phylogeny is the same in all three, with the first split leading to two major clades; clade 1 contains species A, B, and C while clade 2 contains species D, E, and F. A and B are sister species as are E and F. A) Single pair approach consists of all possible pairwise comparisons. Comparisons within each major clade are indicated by solid arrows, comparisons between species of different major clades are indicated by dashed arrows. There are six within-major clade and nine among-major clade comparisons for these six species, N(N-1)/2 overall. B) Hierarchical group analysis involves including all members of a clade within each analysis. The relevant taxa are enclosed in the Venn diagram sets illustrated. Analyses are repeated for all clades. There are N-1 clades for N taxa in a rooted phylogeny; in this example, five clades are

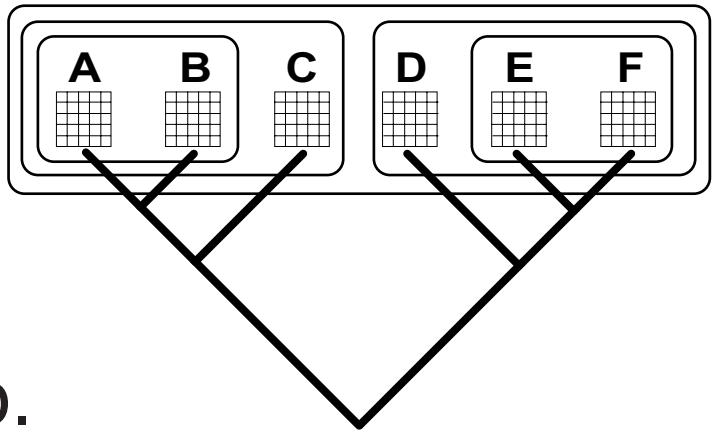
analyzed. C) Illustration of the two primary methods using an ancestral reconstruction approach with phylogenetically independent comparisons. Matrices must be estimated for the hypothetical ancestors G-K from the observed descendent taxa A-F if not observed directly. The dashed arrows indicate comparisons (contrasts) using the phylogenetic independent contrasts method (PIC). In addition to contrasts between extant taxa (A-B, E-F), there are contrasts between ancestors and extant taxa (G-C, I-D) and between estimated ancestors (H-J). All contrasts are between sister taxa. The solid arrows indicate comparisons between ancestors and descendents using the minimum evolution approach (ME). ME allows more explicit partitioning of evolutionary change on the phylogeny but requires greater reliance on the ancestral reconstructions to do so since a greater proportion of comparisons involve ancestors than the PIC method. There are N-1 contrasts and 2(N-1) ME comparisons.

Fig. 3. Contrast between two phylogenies with identical topologies but different branch length distributions. Branch lengths could represent time, amount of molecular evolution, or other characteristics. In this example, the branches are proportional to time, e.g., DNA sequence data evolving under a molecular clock. On the left hand tree, sister species pairs A-B and E-F are very closely related with a long period of shared history (i.e., long branch), while on the right hand tree, the pairs are only slightly more closely related to each other than either member is to the pair's sister taxa, C or D (i.e., short branch, as indicated by arrow). If expected divergence in covariance structure is in any way correlated with time, i.e., any common model of evolution except punctuated equilibrium, then we would expect A and B to be very similar given the left tree but not so given the right. Failure to incorporate branch length information could lead to faulty conclusions about the rate of covariance evolution or about the patterns of similarity as predicted by models being tested.

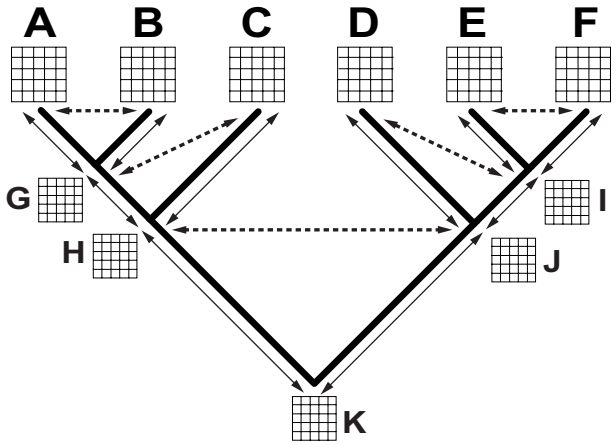

1
2
3
4
5
1
2
3
4
5
Gain 2-4
Loss 2-5
Functional change



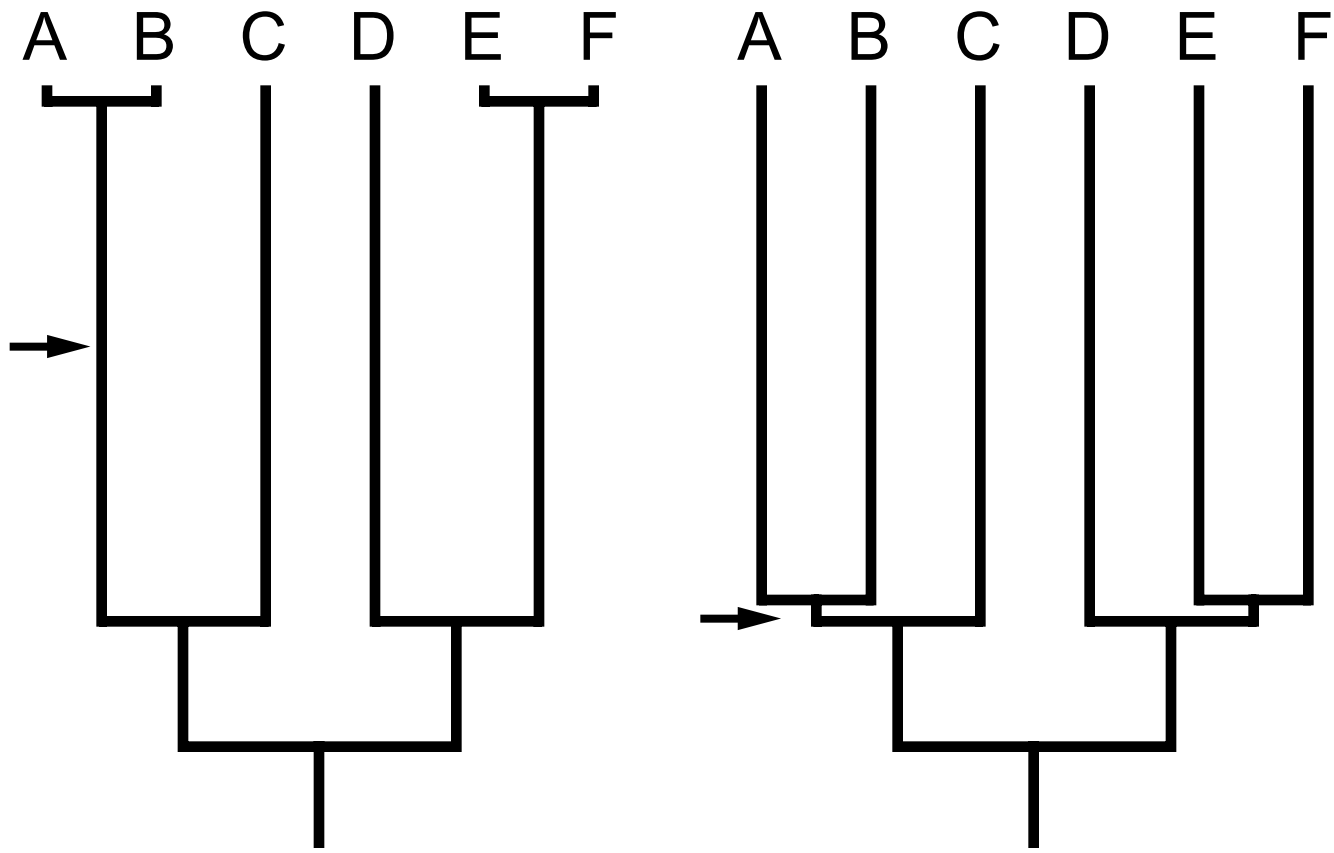
a.



b.



c.


A B C D E F
A B C D E F