EVOLUTION

INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

PUBLISHED BY

THE SOCIETY FOR THE STUDY OF EVOLUTION

Vol. 56 **March 2002** **No. 3**

Evolution, 56(3), 2002, pp. 433–440

INTERPRETATION OF THE RESULTS OF COMMON PRINCIPAL COMPONENTS ANALYSES

DAVID HOULE,[1,2,3] JASON MEZEY,[2] AND PAUL GALPERN[1]

[1]*Department of Zoology, University of Toronto, Toronto, ON M5S 3G5, Canada*

[2]*Department of Biological Science, Florida State University, Tallahassee, Florida 32306-1100*

[3]*E-mail: dhoule@bio.fsu.edu*

Abstract.—Common principal components (CPC) analysis is a new tool for the comparison of phenotypic and genetic variance-covariance matrices. CPC was developed as a method of data summarization, but frequently biologists would like to use the method to detect analogous patterns of trait correlation in multiple populations or species. To investigate the properties of CPC, we simulated data that reflect a set of causal factors. The CPC method performs as expected from a statistical point of view, but often gives results that are contrary to biological intuition. In general, CPC tends to underestimate the degree of structure that matrices share. Differences of trait variances and covariances due to a difference in a single causal factor in two otherwise identically structured datasets often cause CPC to declare the two datasets unrelated. Conversely, CPC could identify datasets as having the same structure when causal factors are different. Reordering of vectors before analysis can aid in the detection of patterns. We urge caution in the biological interpretation of CPC analysis results.

Key words.—Common principal components analysis, Flury hierarchy, matrix comparisons, variance-covariance matrix.

Received June 1, 2001. Accepted November 29, 2001.

Comparisons of two or more covariance matrices are used in many areas of evolutionary biology. Recently, Steppan (1997a,b) examined the macroevolution of phenotypic covariance matrices, Arnold and Phillips (1999) genetic covariance matrix evolution, and Klingenberg and McIntyre (Klingenberg and McIntyre 1998) the developmental origins of fluctuating asymmetry as revealed in covariance matrices. These examples are only some of the most recent (see Roff 2000).

The comparison of two or more variances or variance components is a relatively straightforward statistical procedure for which numerous methods are available (Van Valen 1978; Shaw 1987; Mitchell-Olds and Bergelson 1991). The multivariate analog, the comparison of two or more variance-covariance matrices, is a complex task, and the statistical toolbox for such comparisons is poorly developed. Matrix correlations have been widely used (Lofsvold 1986; Kohn and Atchley 1988), despite the fact that they test the relatively uninteresting null hypothesis that matrices are completely dissimilar (Cowley and Atchley 1992), and have a dubious statistical basis (Shaw 1992). Nevertheless the correlation coefficient is a familiar and therefore heuristic measure of similarity, even if one not well suited to hypothesis testing. Roff et al. (1999) proposed a regression-based version of this approach that shares the problems of matrix correlations, as well as a resampling test for matrix equality. The converse null hypothesis of equal covariance matrices has been tested both with element-by-element comparisons and with maximum-likelihood procedures on whole matrices (Shaw 1991).

In addition to testing the extreme hypotheses of complete dissimilarity or equality, we would like to describe the ways in which matrices are similar or different. Two approaches have been proposed. First, Shaw's (1991) maximum-likelihood method can be used to test more complex hypotheses, although it has rarely been used in this way (Podolsky et al. 1997). Recently, common principal components (CPC) analysis has been suggested for this purpose (Flury 1984, 1988; Phillips and Arnold 1999). The CPC method can be used for the comparison of an arbitrary number of variance-covariance matrices (''covariance matrices'' in our shorthand) and has recently been extended to covariance component matrices (Arnold and Phillips 1999; Phillips and Arnold 1999). The CPC approach has become the most widely used method for comparing phenotypic (Steppan 1997a; Ackermann and Cheverud 2000; Badyaev and Hill 2000; Dodd et al. 2000) and genotypic (Arnold and Phillips 1999; Phillips and Arnold 1999) covariance matrices and for the study of ontogenetic morphological integration (Klingenberg and Zimmermann 1992; Klingenberg and Spence 1993; Klingenberg et al. 1996; Badyaev and Martin 2000).

An integral part of the CPC method is the Flury hierarchy of hypothesis tests, a set of algorithms for determining which eigenvectors differ significantly between covariance matrices (Flury 1988; Phillips and Arnold 1999). For example, using the jump-up algorithm, we accept the null hypothesis of matrix ''equality'' if the matrices cannot be shown to differ. Other levels in the hierarchy are ''proportional,'' in which matrices differ significantly, but a model where all elements are multiplied by a single constant cannot be rejected; ''CPC,'' in which eigenvalues differ from proportional but common eigenvectors are not rejected; and a variety of partial CPC models, in which some eigenvectors can be shown to differ, but others cannot. If the number of traits is p, the ''partial CPC cases'' include matrices sharing from 1 to $p - 2$ common vectors. (Matrices cannot differ in only one vector, because all vectors must be orthogonal. When $p - 1$ vectors have been determined, only one orthogonal direction remains.) In Phillips and Arnold's (1999) terminology these are PCPC($p - 2$), PCPC($p - 3$), . . . PCPC(1). Finally, the results indicate unrelated structure when the model with only one eigenvector in common is rejected. Thus, use of the Flury hierarchy permits the degree of similarity among matrices to be categorized at many levels.

Principal components analysis (PCA), the parent technique on which CPC depends, transforms the data from the space of the original variables, which are correlated, to a set of vectors that are uncorrelated. It captures all of the variation in the original data, while concentrating the variation explained in a few vectors. The forte of PCA is therefore summarization of high-dimensional data. Flury, in developing CPC, similarly emphasized the goal of summarizing multigroup data in as few vectors as possible (Flury 1984, 1987, 1988; Airoldi and Flury 1988; Klingenberg et al. 1996).

This data-summarization function of CPC is not the primary reason that biologists are attracted to the method. We would like to be able to interpret the level of difference found in the Flury hierarchy, rather than simply reduce the number of parameters. Unfortunately, the rationale for such interpretations has never been developed. For example, a particularly important potential application of CPC is in comparing additive genetic covariance matrices (**G**). Constancy of **G** would allow reconstruction of the historical pattern of selection (Lande 1979), which is the rationale for much of the focus on testing the equality of **G** between populations. Arnold and Phillips (1999, p. 1516) argue that CPC is useful when **G** matrices are not equal because it may allow us to address the question of when changes in **G** ''might be so minor that reconstruction of selection is unaffected, or so profound that reconstruction is pointless.'' However, they offer no criteria for linking such questions to the results of a CPC analysis. At present, the appeal of CPC analysis is simply a general conviction that more-detailed hypothesis testing will be more informative than less-detailed testing.

Here, we use simulation to investigate the behavior of the CPC analysis on pairs of matrices that either have equal expectations or that differ in one aspect of their structure. We find that CPC frequently detects no similarity between two matrices that share considerable causal structure. This is an expected, if unfortunate, feature of any analysis based on PCA.

What Are Matrices Made of?

To generate simulated covariance matrices, we necessarily assume models for the causes of covariance structure. Our conception of variation is that it results from a complex set of causal factors. Each of these causal factors leaves a characteristic pattern of covariation among the units being analyzed. Some of these causes affect only one trait, including most forms of measurement error. Other sources of variation may cause whole suites of traits to covary. We assume that variation in each trait is determined by one or more underlying causal factors. Causal factors affect one or more traits, and more than one factor can affect the same trait. Factors that affect multiple traits lead to covariance structure in a sample of individuals. If a pair of factors affects at least one trait in common, we say that the factors overlap.

The model can be represented algebraically as follows. Any p traits may be affected by any q factors. The $p \times 1$ vector of trait values (**t**) is determined by $q \times 1$ vector of scores on each factor (**s**) and the $p \times q$ factor/trait loadings matrix (**F**), plus a $p \times 1$ vector of deviations (**d**) to simulate trait-specific variation and measurement error:

$$\mathbf{t} = \mathbf{Fs} + \mathbf{d} \tag{1}$$

Both **s** and **d** are assumed to be normally distributed with a mean of zero and no covariances between elements. The covariance structure of a dataset is determined by the **F** matrix of factor loadings.

The reader may recognize that this is a conception of matrix structure that bears no relation to that embodied in PCA. Two features of PCA are responsible for the difference. First, the vectors identified in PCA are constrained to be orthogonal to each other, whereas causal factors need not have orthogonal effects on the phenotype. Second, PCA may be thought of as sequentially determining the location of vectors in such a way as to maximize the variance explained by each successive vector. Thus, the first vector explains the most variance, the second the second most, etcetera. The causal factors themselves are not constrained to lie in directions of maximal variance.

We have chosen this representation of causation of matrix structure because it is one that seems both simple and biologically reasonable. We can readily imagine more complex models of causation, for example, nonlinear relationships between factor scores and phenotypic effects, or interactions between different causal factors. We have difficulty imagining that the causes of variation in real organisms act orthogonally. The results of our simulations follow directly from this assumption about causation.

To take a hypothetical example, morphometric variation might reflect variation in $q = 2$ factors. The first factor represents genetic variation in a developmental pathway that affects the relative lengths of limbs and indirectly affects the overall size of the organism through its impact on feeding efficiency. The second factor might represent variation in developmental temperature and affect limb dimensions as well as the overall shape and size of the body. Both temperature and genetic variation in development affect overall size and shape, but in slightly different ways. The result is vectors of effects that are not orthogonal. PCA would spread

TABLE 1. Parameter values used to create simulated data from equation (1). Population is indexed by the superscript. **S** is a vector containing the variances of the corresponding factor (the transposed diagonal of the factor variance-covariance matrix of **s**). **D** is a vector containing the trait-specific variances (the transposed diagonal of the variance-covariance matrix of **d**). Except in cases 1.1 and 1.2 in population 2, **D** = [0.25 0.25 0.25 0.25 0.25 0.25]. Bold entries reflect elements in which populations 1 and 2 differ.

Case:	1.1,† 1.2‡	2.1	2.2	3.1	3.2	3.3	3.4	4.1	4.2
Population 1:									
$\mathbf{F}^1$	$\begin{bmatrix}1&0&0\\1&0&0\\0&1&0\\0&1&0\\0&0&1\\0&0&1\end{bmatrix}$	$\begin{bmatrix}1&0&0\\1&0&0\\0&1&0\\0&1&0\\0&0&1\\0&0&1\end{bmatrix}$	$\begin{bmatrix}1&0&0\\1&1&0\\1&1&1\\0&1&1\\0&0&1\\0&0&0\end{bmatrix}$	$\begin{bmatrix}1&0\\1&0\\0&1\\0&1\\0&0\\0&0\end{bmatrix}$	$\begin{bmatrix}1&0\\1&0\\0&1\\0&1\\0&0\\0&0\end{bmatrix}$	$\begin{bmatrix}1&0\\1&0\\0&0\\0&1\\0&0\\0&0\end{bmatrix}$	$\begin{bmatrix}1&0\\0&1\\0&0\\0&0\\0&0\\0&0\end{bmatrix}$	$\begin{bmatrix}1&1&1&1&1&0\\1&1&1&1&0&1\\1&1&1&0&1&1\\1&1&0&1&1&1\\1&0&1&1&1&1\\0&1&1&1&1&1\end{bmatrix}$	$\begin{bmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&1&0&0&0\\0&0&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\end{bmatrix}$
$\mathbf{S}^1$	[3 2 1]	[3 2 1]	[3 2 1]	[3 2]	[3 2]	[4 1]	[3 3]	[1 1 1 1 1 1]	[6 5 4 3 2 1]
Population 2:									
$\mathbf{F}^2$	$\begin{bmatrix}1&0&0\\1&0&0\\0&1&0\\0&1&0\\0&0&1\\0&0&1\end{bmatrix}$	$\begin{bmatrix}1&0&0\\1&0&0\\0&1&0\\0&1&0\\0&0&1\\0&0&1\end{bmatrix}$	$\begin{bmatrix}1&0&0\\1&1&0\\1&1&1\\0&1&1\\0&0&1\\0&0&0\end{bmatrix}$	$\begin{bmatrix}1&0\\1&0\\0&1\\0&1\\0&0\\0&0\end{bmatrix}$ to $\begin{bmatrix}1&0\\1&0\\0&\mathbf{0}\\0&\mathbf{0}\\0&\mathbf{1}\\0&\mathbf{1}\end{bmatrix}$	$\begin{bmatrix}1&0\\1&0\\0&1\\0&1\\0&0\\0&0\end{bmatrix}$ to $\begin{bmatrix}1&\mathbf{1}\\1&\mathbf{1}\\0&\mathbf{0}\\0&\mathbf{0}\\0&0\\0&0\end{bmatrix}$	$\begin{bmatrix}1&0\\1&0\\\mathbf{1}&0\\0&1\\0&0\\0&0\end{bmatrix}$	$\begin{bmatrix}1&\mathbf{1}\\\mathbf{1}&\mathbf{-1}\\0&0\\0&0\\0&0\\0&0\end{bmatrix}$	$\begin{bmatrix}1&1&1&1&1&0\\1&1&1&1&0&1\\1&1&1&0&1&1\\1&1&0&1&1&1\\1&0&1&1&1&1\\0&1&1&1&1&1\end{bmatrix}$	$\begin{bmatrix}1&0&0&0&0&0&\mathbf{0}\\0&1&0&0&0&0&\mathbf{0}\\0&0&1&0&0&0&\mathbf{1}\\0&0&0&1&0&0&\mathbf{1}\\0&0&0&0&1&0&\mathbf{0}\\0&0&0&0&0&1&\mathbf{0}\end{bmatrix}$
$\mathbf{S}^2$	[3 2 1]	[3***k*** 2 1]	[3***k*** 2 1]	[3 2]	[3 2]	[4 1]	[**1.5 1.5**]	[***k*** 1 1 1 1 1]	[6 5 4 3 2 1 **3**]

† Case 1.1: $\mathbf{D}^2 = [0.25\boldsymbol{k}\ \ 0.25\ \ 0.25\ \ 0.25\ \ 0.25\ \ 0.25]$.
‡ Case 1.2: $\mathbf{D}^2 = \boldsymbol{k}\mathbf{D}^1$.

the effects of each cause over multiple orthogonal vectors—first into a ''size'' vector that all traits load positively onto, and then onto other vectors that express changes in the relative sizes of parts. The interpretational difficulty that arises from this situation is that a change in only one of the two causal vectors, for example, a change in the variance in temperature that individuals experience during development, affects all of the eigenvectors in a PCA. We present examples of this effect in our simulations below.

METHODS

To investigate the behavior of CPC analysis, we simulated phenotypic covariance matrices with a known underlying causal structure corresponding to the model in equation (1). We restricted our attention to phenotypic matrices because we are interested in the interpretation of matrix structure that CPC potentially allows. Comparisons of variance component matrices, such as **G**, pose additional uncertainties and difficulties (Phillips and Arnold 1999) that we do not address.

We analyzed variation in two simulated populations that differed in a single parameter of the causal model. For each set of parameter combinations, 100 datasets were created, each with a sample size of 300 individuals, unless otherwise noted. The parameters used for each simulation are summarized in Table 1. The top section of Table 1 gives the parameter values in population 1; the lower section gives those in population 2. Results of most cases considered are shown in Figure 1. In some of the cases, we obtained some parameters in population 2 by multiplying those in population 1 by a factor k that we varied from one to some larger value. The x-axis of the corresponding panels in Figure 1 express CPC results as k varies. CPC analyses were performed with software written by Phillips (1998).

We investigated the use of two criteria for choosing the best-fitting level of similarity in the CPC, the jump-up and Akaike information criterion (AIC). Phillips and Arnold (1999) argue that the most useful indicator of matrix relationships is found by means of the jump-up approach, because it is least sensitive to small differences in covariance structure. In the jump-up approach, an alternative hypothesis of unrelated structure is tested against a null hypothesis from each level of the Flury hierarchy. If the test at a lower level is nonsignificant, the next higher level is tested. If a test is significant, indicating a departure from shared covariance structure at that level, the next lowest level provides the best explanation of matrix relationships. Flury (1988, pp. 151–152) favored the use of the ''model-fitting'' approach, where the best-fitting model is chosen according to the AIC. This approach maximizes the variance in the data explained for each parameter fit.

By default, the Phillips implementation of CPC analysis constrains the two matrices to share the vectors with the largest eigenvalues from a common model fit to the pooled data. It also allows the user to change the order in which the shared vectors are considered. In a few cases we explored the consequences of this reordering of vectors.

Testing the Common Principal Components Method

We contrast two sets of criteria for judging the performance of the CPC method. From the statistical point of view, the outcome of CPC analysis is determined by the known differences in eigenstructure of our pairs of simulated matrices

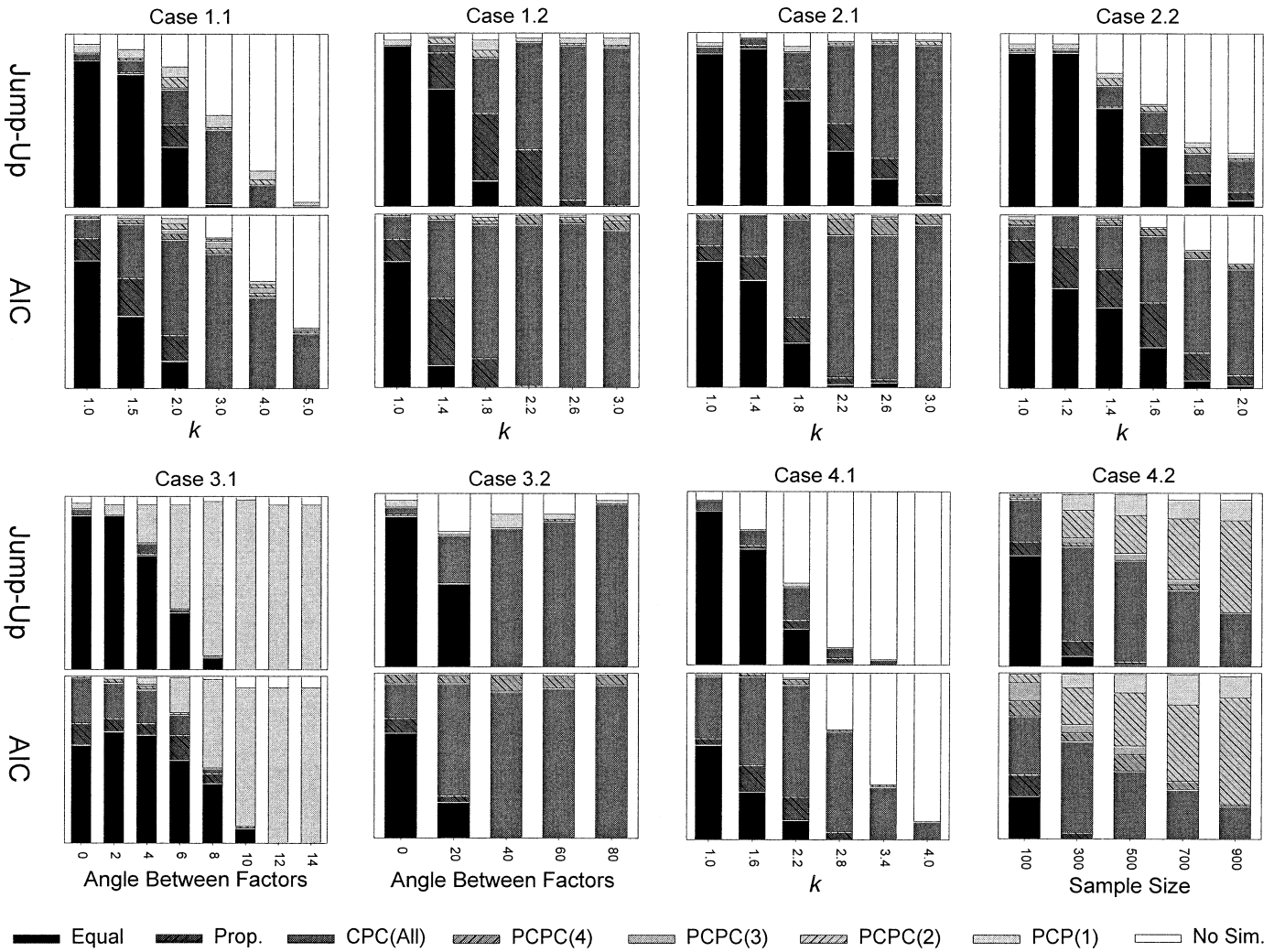

FIG. 1. Results of common principal components (CPC) analyses of pairs of covariance matrices that differ in their structure. Each bar gives the proportion of results in each of the eight categories of the Flury hierarchy for six traits (equality, proportionality, CPC[all], CPC[4], CPC[3], CPC[2], CPC[1], and no similarity; see introduction for explanation). Covariance matrices were generated according to equation (1) with the parameter values in Table 1. In some cases, parameters in population 1 were multiplied by a factor k to yield those in population 2, and the results are graphed as a function of k. Upper panels show the results of a jump-up analysis; lower panels show the results with Akaike information criterion (AIC).

and by the statistical power of the tests. The result corresponding to the known differences in structure, leaving aside power issues, is referred to as the ''expected result.'' Alternatively, knowing what differences exist in the causal structure of the two matrices, we can ask whether CPC reflects the changes in this structure in an interpretable way. We refer to this as the ''biological'' point of view. The discrepancy between the expected and biological interpretations is the major theme of our results.

Case 1: Differences in error variance

In these cases, we considered two populations that shared an orthogonal causal structure to their trait covariances but differed in their error or trait specific variances. Biologically, this case could correspond to one in which population samples are analyzed at different times or by different observers or in which trait-specific variances differed for genetic or environmental reasons. In these cases we assumed a causal structure with three factors, where two different traits load on each factor, as shown in the **F** matrix in Table 1.

Case 1.1. Trait-specific variance of a single trait differs.—In this set of simulations, the trait-specific variance associated with trait 1 in population 2 was k times as large as that in population 1, where k was varied between 1 and 5. When trait-specific variances are equal, the expectations of the matrices compared are equal. As shown in Figure 1, the CPC consistently returned equality in this case. However, when $k > 1$, the expected result is no similarity, because the added variance in one trait affects the directions of all the resulting eigenvectors. Biologically, we might hope to find a method that ignores the trait-specific differences and recognizes the common structure of the two datasets. Alternatively, we might hope that the added variance would affect only one vector and still allow the recognition of shared structure. As

expected from statistical reasoning, CPC tended to return a result of no similarity when the difference in error variance was large, rather than limiting the differences to a few vectors, as the naive biological expectation suggests. The jump-up approach appears to have higher power to detect the expected outcome. Note that at intermediate k values CPC tended to return CPC(all) and only rarely returned intermediate-level CPC models. This is a common result. When the statistically expected result was no similarity, but the differences between the populations were subtle, the outcome was often CPC(all). With larger differences, CPC jumped to the expected result of no similarity.

Case 1.2. Error variance associated with all traits.—In this set of simulations, the trait-specific variance for all traits in population 2 was multiplied by a factor k between 1 and 3. When $k > 1$, the expected result becomes CPC(all), as eigenvalues but not eigenvectors are affected. Both AIC and jump-up methods returned CPC(all) with high probability when k was large, but AIC appeared to have greater power when k was modest. In this case, biological and expected statistical outcomes coincided.

Case 2: Changes of factor variance

In these cases, we assumed the same three causal factors acted in each population, but the variance associated with the first factor (the factor with the most variance) differed between the two populations by a factor k. Differences in causal variances arise if the populations differ in genetic or environmental variance of factors, or if one population is more canalized than another.

Case 2.1. Orthogonal factors.—If causal factors are orthogonal, changing the variance of a factor leaves the principal component orientation unchanged, and the expected result is therefore CPC(all). Both jump-up and AIC converged on the expected result as the differences between the populations increased, but AIC converged faster. The factor k must be fairly large for the CPC method to report the expected answer at an appreciable rate. In this case, the naive biological interpretation of the CPC result is correct. Examination of the eigenvalues associated with each component correctly revealed the source of the departure from matrix equality.

Case 2.2. Nonorthogonal factors.—When factors are not orthogonal and factor variances differ, the expected statistical result is no similarity. The jump-up test converged on the expected result much faster than AIC as k was increased, but when k was modest, CPC(all) was the most common result. From the biological perspective, these outcomes are disappointing. The factor structures of the populations were identical, so the CPC result of no similarity did not match biological reality. The result CPC(all) appears to match biological reality, but this agreement with biological interpretation is fortuitous. The CPC(all) result was returned only because of a lack of statistical power.

Case 3: Different factor orientation

In these cases, the two populations differed in the orientation of one of the two factors. Biologically, this difference might correspond to changes in the regulation of developmental pathways involved in trait construction.

Case 3.1. Orthogonal factors.—In this case, the factors were always orthogonal within populations, but the orientation of factor 2 in population 2 was rotated between 0° and 90° relative to factor 2 in population 1. Vectors at these two extreme angles are shown in Table 1; between 0° and 90° traits 3 to 6 all load on factor 2, with the additional constraint that traits 3 and 4 and traits 5 and 6 always have identical loadings. For example, we used $[0, 0, 0.866, 0.866, 0.5, 0.5]^T$ as factor 2 in population 2 (the second column of the $\mathbf{F}^2$ matrix) to generate an angle of 30° relative to factor 2 in population 1. The expected result changes from equality when the factors are equal, to PCPC(1) when they are not. CPC is extremely sensitive to small changes in angle and readily returns the expected result. As in case 2.1, the naive biological interpretation is correct when factors within populations are orthogonal. Jump-up performed better than AIC in this case.

Case 3.2. Nonorthogonal factors.—In this case, the factor structure in population 1 was orthogonal, whereas factor 2 in population 2 was rotated from 90° (orthogonal to factor 1) to 0° (identical to factor 1) relative to factor 1 in an analogous manner to the rotation in case 3.1. At 90° the expected result is equality, at 0° the expected result is CPC(all). At intermediate angles, however, the expected result is no similarity. The no-similarity result was returned with low frequency by the jump-up approach between 20° and 80°, and was essentially never returned by the AIC approach. Above 60°, most of the results were equality, and below 40°, most were CPC(all). This case is problematic for the CPC method. The expected result was only returned at a low rate, even for intermediate angles, where the differences would be easiest to identify. Biologically, these two populations had a single factor in common, and the statistically expected result was clearly in conflict with this. With the sample sizes considered here, CPC tended to miss the differences between populations altogether.

Case 3.3. Different orientation of the first factor.—In this case, factors in both populations were orthogonal, but the populations differed in factor 1. This case was therefore similar to case 3.1, except that in this case it was the primary factor that differed between the two populations. Clearly, the statistically and biologically expected result was PCPC(1), but the similarity was in the vector with the second largest eigenvalue.

Without reordering, both jump-up and AIC tended to return no similarity, and even when PCPC(1) resulted, the eigenvector found to be in common did not correspond to the shared factor; it actually represented a composite of the different first factors in the two populations. When the reordering option was used, however, both methods recovered the expected result PCPC(1) at a rate greater than 90% for sample sizes of 50 or 300 (results not shown).

Case 3.4. All factors are different.—In this case, we constructed an example in which the causal factors differ between the two populations, but the covariance matrices they generated were equal. The CPC method returned the expected statistical result of equality for these two populations at a rate greater than 95% by the jump-up method (results not shown). This case illustrates that, even when PCs are found

to be in common, they may not correspond to common biological factors.

Case 4: Overdetermined system

In our last two cases, we assumed there are at least as many factors as traits. A large number of genetic loci are capable of influencing phenotypes, and each may have a unique pattern of effects. Systems with more causal factors than characters may therefore be biologically realistic.

Case 4.1. Factors affect most traits.—One interpretation of the size factor often found in morphological datasets is that many aspects of variation affect the overall growth of organisms, although each causal factor does so in a slightly different way. To model this situation, we assumed that each population was affected by six factors, each affecting all but one of the six traits. In population 2, the variance in factor 1 was multiplied by k between 1 and 4. When $k > 1$, the expected result is no similarity, and the jump-up method correctly returned this answer when k was large. However, from a biological point of view, we would like to know that these populations share a considerable number of causal factors. The CPC method seems particularly unlikely to indicate shared causal structure when factors overlap a great deal. When the difference in variance was not great, AIC tended to return CPC(all), which can be interpreted as shared factor structure, the correct biological interpretation, although again only because of lack of statistical power, as in case 2.2.

Case 4.2. One population has an extra factor.—In this case, population 1 had six factors, and each affected only one trait. Population 2 had these same six factors, plus a seventh factor that affected traits 3 and 4. With no reordering, both jump-up and AIC converged on the expected result PCPC(2) as sample size increased (not shown), although for smaller sample sizes, CPC(all) is the most frequent finding. Factor 7 in population 2 obscured identification of further similarity beyond PCPC(2). The last panel in Figure 1 shows the results when manual reordering was used to solve this problem. With reordering, CPC converged on the expected result of PCPC(4) as sample size increased.

Discussion

The CPC method is an unquestioned advance over previous methods of covariance matrix comparison. Under the assumption of multivariate normality, it allows a simple test for matrix equality, proportionality, or common structure. It also allows matrix structure to be decomposed and intermediate levels of similarity to be assessed. Our results suggest, however, that these intermediate levels of similarity are unlikely to have a useful biological interpretation. We discuss the implications of our results first for the general use of the method then for biological interpretation of the results.

General Use of Common Principal Components Analysis

Our results provide some useful general indications of the behavior of the CPC method. Even under the ideal conditions of our simulations, the power of CPC tests is limited. Large sample sizes were necessary for CPC to return the expected principal components model at an appreciable rate. For example, in case 4.2, even though the expected statistical result is PCPC(4), CPC returns a considerable number of results as CPC(all) or less difference when sample sizes are less than 700. These results suggest that differences between covariance matrices are difficult to interpret. For example, Arnold and Phillips (1999) found that comparisons of **P** matrices (using individual estimates) between sexes and between two populations resulted in PCPC(2)–PCPC(4), depending on the comparison. Comparisons of covariance-component-based **G** matrices returned CPC(all). This difference may simply follow from the higher sampling variance of **G** matrices than of **P** matrices. Examination of confidence limits on eigenvalues and eigenvectors (e.g., Airoldi and Flury 1988) would be helpful in interpreting failures to reject particular null hypotheses.

The Flury hierarchy specifies a restricted set of hypotheses to test, thus simplifying the CPC analysis. The disadvantage of this simplification is that a finding of unrelated matrices means only that the lowest-level model tested can be rejected, not that all models of similarity have been rejected. Differences in vectors associated with large eigenvalues often prevent recognition of shared structure lower in the hierarchy, as shown in cases 3.3 and 4.2. Manual reordering can help to reveal shared vectors lower in the hierarchy. Cases in which manual reordering is helpful can be identified from a search for similar eigenvectors in the full CPC model and the PCA analyses on each population. However, reordering may not prove to be helpful in more realistic cases with nonorthogonal, overlapping causal factors. In general, a change in any one causal vector can alter all of the resulting eigenvectors detected by PCA, as shown in many of our simulations. The cases we used to demonstrate reordering were both structured to prevent this problem.

These simulations also indicate differences in outcome between the jump-up and AIC methods for determining the proper level in the Flury hierarchy. When covariance structures are equal, the jump-up approach returns a greater proportion of matrix equality than AIC. When the expected large-sample statistical result is no similarity, but the differences are fairly subtle, jump-up also returns a higher proportion of expected results than AIC. However, when the expected statistical result is an intermediate level of shared structure, such as CPC(all) or a partial CPC model, AIC returns a greater proportion of expected results. These differences in performance disappear as sample sizes and power become large, but for the sample sizes likely to be employed in biological studies, this result should be of some concern.

When two matrices with equal expectations were compared in our simulations, the jump-up method usually rejected equality at a rate closer to 10% than the desired 5%. This result is expected because the jump-up approach requires that a large number of null hypotheses must be accepted to reach the highest levels of similarity. A more conservative P-value could be chosen to compensate for this.

Biological Interpretation of Common Principal Components Results

CPC analysis is based on PCA and therefore shares its advantages and disadvantages. CPC was originally proposed

as a method for summarizing the variation in two or more matrices simultaneously (e.g., Airoldi and Flury 1988; Klingenberg et al. 1996). We see no problems in its use in this context.

However, some evolutionary biologists are turning to CPC with the more complex goal of diagnosing and understanding the nature of the changes that underlie a difference between covariance matrices. For example, Phillips and Arnold (1999, p. 1507) note that the ''structure of **G** . . . may reflect the underlying pattern of developmental interactions and associations'' and present the CPC method as a tool for investigating such structure. As discussed above, we expect that the actual causes of covariance structure in real populations are not likely to correspond to the notions of structure embodied in PCA or CPC analysis. Conversely, different causal relationships may result in the appearance of similarity, as in case 3.4. It is clear, therefore, that no simple relationship exists between the level of relationship found in a CPC analysis and other conceptions of matrix similarity. CPC tends to spread any differences over many of the vectors it extracts and often over all of them.

When testing for differences among covariance patterns between populations, we would like the results to reflect our notions of the causes of covariance structure. For example, we might expect that a difference between two populations in the nature of one causal factor would result in the detection of one vector as different between populations. We might expect that if there are only quantitative changes in the amount of variation contributed by each causal factor, that the resulting matrices would still retain the same set of component vectors, resulting in a CPC(all) result. Finally, we might expect that uncorrelated differences in any one trait would not affect the overall pattern of vectors uncovered in the analysis. The chief value of the simulations reported here is to disabuse biologists of these simple expectations. Neither the CPC nor the departures from a CPC model seem likely to be informative. Inferences from CPC about biological causation are likely to be wrong.

If our conception of the cause of matrix structure is so alien to that of CPC, is there an alternative notion of causation that is more appropriate to use? We do not know of any. We can readily conceive of models for covariance structures that are more complex—for example, those with nonlinear relationships between causes and effects or in which different causal factors interact—but we do not believe that causal factors have orthogonal effects on the phenotype. Thus, the notion of causation that we believe to be biologically reasonable is, a priori, inconsistent with the procedure of CPC.

When causal factors are not orthogonal, single changes in the causal structure of covariation often lead to a diagnosis that two matrices are completely dissimilar. This tendency is increased by the default inclusion of the first eigenvector in all hypotheses tested with the Flury hierarchy. Thus, a difference in the first eigenvector of two matrices will cause a finding of unrelated matrices, regardless of any other similarities between them. The performance of the CPC method in recent studies is consistent with the presence of such effects. For example, Steppan's (1997a) CPC analysis found that populations within the same subspecies sometimes had very similar covariance matrices (CPC[all]), although in other cases they showed no similarity. CPC comparisons among species usually resulted in a finding of no similarity. In contrast, matrix correlations at these two levels are extremely high, ranging from an average of 0.96 among different populations within the same subspecies to 0.93 among species groups. Steppan's (1997a) CPC results are thus consistent with simulations showing that relatively restricted changes in causal structure can result in a finding of no similarity. Another interpretation of these results is that power is only sufficient to detect differences in the first few eigenvectors, so effectively no tests are being made of the intermediate hypotheses.

The model of covariance matrix structure that we have adopted here is formally equivalent to that of factor analysis, suggesting that factor analysis may be more suitable for biologically interpretable matrix comparisons. Usually factor analysis is carried out with no a priori constraints on the nature of the vectors that underlie the data, in which case the analysis is called ''exploratory factor analysis.'' A few evolutionary biologists have used ad hoc exploratory factor analysis to compare covariance matrices (Gale and Eaves 1972; Arnold 1981; Goodin and Johnson 1992; Paulsen and Nijhout 1993). Unfortunately, exploratory factor analysis assumes that the covariance matrix can be summarized by a small number of factors, much less than the number of traits, and that each cause only affects a subset of traits. Both of these simplifying assumptions are likely to be wrong.

Confirmatory factor analysis, in which hypotheses about the nature of covariance structure are formulated prior to analysis, is more promising. Confirmatory analyses have been applied to two morphometric datasets (Zelditch 1988; Zelditch and Carmichael 1989) to test simple models concerning the nature of development. A factor-analytic analog of CPC analysis has been outlined by statisticians (Jöreskog 1971; Tisak and Meredith 1989) but has not to our knowledge been used for biological analyses. More generally, structural equation modeling, which includes confirmatory factor analysis as a special case, can be used to accommodate a wider variety of prior information and certain kinds of interactions between causal factors (Kline 1998). Development and experimental verification of causal models for covariance structure, coupled with these modeling techniques, may prove a fruitful approach to matrix comparisons in the future.

Acknowledgments

We thank P. C. Phillips and S. J. Steppan for their advice in applying CPC analysis and their comments on the manuscript. We also thank D. Jackson, T. Kawecki, C. P. Klingenberg, L. Rowe, D. Currie, P. Peres-Neto, I. Dworkin, J. A. Morrow, and reviewers for their comments. This work was supported in part by a National Science and Engineering Research Council scholarship to P. Galpern, an NSERC grant to D. Houle, and by Florida State University.

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.

Airoldi, J. P., and B. D. Flury. 1988. An application of common

principal component analysis to cranial morphometry of *Microtus californicus* and *M. ochrogaster* (Mammalia, Rodentia). J. Zool. 216:21–36.
Arnold, S. J. 1981. Behavioral variation in natural populations. I. Phenotypic, genetic and environmental correlations between chemoreceptive responses to prey in the garter snake, *Thamnophis elegans*. Evolution 35:489–509.
Arnold, S. J., and P. C. Phillips. 1999. Hierarchical comparison of genetic variance-covariance matrices. II. Coastal-inland divergence in the garter snake, *Thamnophis elegans*. Evolution 53: 1516–1527.
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.
Badyaev, A. V., and T. E. Martin. 2000. Individual variation in growth trajectories: phenotypic and genetic correlations in ontogeny of the house finch (*Carpodacus mexicanus*). J. Evol. Biol. 13:290–301.
Cowley, D. E., and W. R. Atchley. 1992. Comparison of quantitative genetic parameters. Evolution 46:1965–1967.
Dodd, R. S., Z. A. Rafii, and A. Bousquet-Mélou. 2000. Evolutionary divergence in the pan-Atlantic magrove *Avicennia germinans*. New Phytol. 145:115–125.
Flury, B. D. 1984. Common principal components in *k* groups. J. Am. Stat. Assoc. 79:892–898.
———. 1987. Two generalizations of the common principal component model. Biometrika 74:59–69.
———. 1988. Common principal components and related multivariate models. Wiley, New York.
Gale, J. S., and L. J. Eaves. 1972. Variation in wild populations of *Papaver dubium*. V. The application of factor analysis to the study of variation. Heredity 29:135–149.
Goodin, J. T., and M. S. Johnson. 1992. Patterns of morphological covariation in *Partula*. Syst. Biol. 41:292–304.
Jöreskog, K. G. 1971. Simultaneous factor analysis in several populations. Psychometrika 36:409–426.
Kline, R. B. 1998. Principles and practice of structural equation modeling. Guilford Press, New York.
Klingenberg, C. P., and G. S. McIntyre. 1998. Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with procustes methods. Evolution 52: 1363–1375.
Klingenberg, C. P., and J. R. Spence. 1993. Heterochrony and allometry: lessons from the waterstrider genus *Limnoporus*. Evolution 47:1834–1853.
Klingenberg, C. P., and M. Zimmermann. 1992. Static, ontogentic, and evolutionary allometry: a multivariate comparison in nine species of water striders. Am. Nat. 140:601–620.
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.
Kohn, L. A. P., and W. R. Atchley. 1988. How similar are genetic correlation structures? Data from mice and rats. Evolution 42: 467–481.
Lande, R. 1979. Quantitative genetic analysis of multivariate evolution applied to brain:body size allometry. Evolution 33: 402–416.
Lofsvold, D. 1986. Quantitative genetics of morphological differentiation in *Peromyscus*. I. Tests of the homogeneity of genetic covariance structure among species and subspecies. Evolution 40:559–573.
Mitchell-Olds, T., and J. Bergelson. 1991. Statistical genetics of an annual plant, *Impatiens capensis*. I. Genetic basis of quantitative variation. Genetics 124:407–415.
Paulsen, S. M., and H. F. Nijhout. 1993. Phenotypic correlation structure among elements of the color pattern in *Precis coenia* (Lepidoptera: Nymphalidae). Evolution 47:593–618.
Phillips, P. C. 1998. CPC: Common principal components analysis. Available via http://www.uoregon.edu/~pphil/programs/cpc/cpc.htm.
Phillips, P. C., and S. J. Arnold. 1999. Hierarchical comparison of genetic variance-covariance matrices. I. Using the Flury hierarchy. Evolution 53:1506–1515.
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.
Roff, D. 2000. The evolution of the G matrix: selection or drift? Heredity 84:135–142.
Roff, D. A., R. 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.
Shaw, R. G. 1987. Maximum-likelihood approaches applied to quantitative genetics of natural populations. Evolution 41: 812–826.
———. 1991. The comparison of quantitative genetic parameters between populations. Evolution 45:143–151.
———. 1992. Comparison of quantitative genetic parameters: reply to Cowley and Atchley. Evolution 46:1967–1969.
Steppan, S. J. 1997a. Phylogenetic analysis of phenotypic covariance structure. I. Contrasting results from matrix correlations and common principal component analyses. Evolution 51: 571–586.
———. 1997b. Phylogenetic analysis of phenotypic covariance structure. II. Reconstructing matrix evolution. Evolution 51: 587–594.
Tisak, J., and W. Meredith. 1989. Exploratory longitudinal factor analysis in multiple populations. Psychometrika 54:261–281.
Van Valen, L. 1978. The statistics of variation. Evol. Theory 4: 33–43.
Zelditch, M. L. 1988. Ontogenetic variation in patterns of phenotypic integration in the laboratory rat. Evolution 42:28–41.
Zelditch, M. L., and A. C. C. Carmichael. 1989. Ontogenetic patterns of developmental and functional integration in skulls of *Sigmodon fulviventer*. Evolution 43:814–824.

Corresponding Editor: T. Kawecki