

METHODS FOR STUDYING MORPHOLOGICAL INTEGRATION AND MODULARITY

ANJALI GOSWAMI

Department of Genetics, Evolution and Environment and Department of Earth Sciences, University College London, Wolfson House 408, 4 Stephenson Way, London NW1 2HE UK

and

P. DAVID POLLY

Department of Geological Sciences, Indiana University, 1001 E. 10th Street, Bloomington, IN 47401 USA

ABSTRACT.—Morphological integration and modularity are closely related concepts about how different traits of an organism are correlated. Integration is the overall pattern of intercorrelation; modularity is the partitioning of integration into evolutionarily or developmentally independent blocks of traits. Modularity and integration are usually studied using quantitative phenotypic data, which can be obtained either from extant or fossil organisms. Many methods are now available to study integration and modularity, all of which involve the analysis of patterns found in trait correlation or covariance matrices. We review matrix correlation, random skewers, fluctuating asymmetry, cluster analysis, Euclidean distance matrix analysis (EDMA), graphical modelling, two-block partial least squares, RV coefficients, and theoretical matrix modelling and discuss their similarities and differences. We also review different coefficients that are used to measure correlations. We apply all the methods to cranial landmark data from an ontogenetic series of Japanese macaques, *Macaca fuscata* to illustrate the methods and their individual strengths and weaknesses. We conclude that the exploratory approaches (cluster analyses of various sorts) were less informative and less consistent with one another than were the results of model testing or comparative approaches. Nevertheless, we found that competing models of modularity and integration are often similar enough that they are not statistically distinguishable; we expect, therefore, that several models will often be significantly correlated with observed data.

INTRODUCTION

“In nature we never see anything isolated, but everything in connection with something else which is before it, beside it, under it and over it.”

Johann Wolfgang Von Goethe

THAT ORGANISMS are composed of multiple parts with varying levels of interrelationships based on shared function or developmental history is intuitive. The study of the phenotypic expression of these relationships, their causes, and their ultimate integration into a cohesive whole organism has blossomed into a rich field of study since the publication of Olson and Miller’s groundbreaking book, *Morphological Integration* (Olson and Miller, 1958). More recently, modularity – the idea that integration occurs within

semi-autonomous units, or modules, with very little integration across modules – has subsumed studies of morphological integration. The significance of modular organisation is that modules can vary and evolve independently of each other while still maintaining the integrity of the functional or developmental unit.

The recent emergence of modularity is of course not the first time **independent evolutionary morphological units** have been proposed; the approach of examining functional or developmental units separately has a lengthy history and has provided an important framework for studying morphological evolution (e.g., van der Klaauw, 1948-1952; Moss and Young, 1960; Schwenk, 2001). Modularity expands on past approaches in being equally applicable to genetic, developmental, and morphological aspects of an organism or organisms, from signaling pathways to colonies

(Schlosser and Wagner, 2004). Examining modularity at all of these scales, and then using it to study the relationships among these scales and across diverse biological systems has proven to be an exciting and inclusive area of evolutionary study. Indeed, identifying the biological drivers of morphological integration was central to the methodology originally proposed by Olson and Miller (1958). They used trait correlations to define ρ -groups, quantitatively-derived clusters in which all trait pairs are highly correlated, at some specified but arbitrary level. F -groups, in contrast, were originally qualitatively defined as being functionally or developmentally integrated, although this definition was later expanded to include any biological relationship. In a series of empirical analyses, Olson and Miller (1958) showed a strong correspondence between ρ -groups and F -groups, thereby verifying a central point: quantitative analysis of trait correlations can accurately recover biologically real components. This result is reflected in the redefinition of morphological integration by Chernoff and Magwene (1999) as “the correspondence of patterns of covariation among traits to *a priori* or *a posteriori* biological hypotheses”.

The fact that integration and modularity can be studied with only morphological data is of interest to comparative morphologists and paleontologists because the quantitative approaches that we describe below can be used to test hypotheses of developmental integration and modularity in extinct or rare taxa, contrasting many other developmental or genetic questions which do not easily lend themselves to testing with paleontological data. More importantly, identifying phenotypic modules in both fossil and recent species may have important implications for understanding morphological evolution, as traits do not vary and evolve independently of each other, although they are often treated as independent in large-scale studies, such as phylogenetic analyses (P.J. Wagner, 1998; O’Keefe and Wagner, 2001; González-José et al., 2008; Sadleir and Makovicky, 2008; Goswami and Polly, 2010a). The relationships among traits can influence their evolution in multiple ways: integration may constrain variation, modularity may circumvent the effects of developmental canalization and genetic pleiotropy. For these reasons, many workers have tied integration and modularity to fundamental questions about evolutionary rates, evolvability, and the generation of new morphologies (Olson and Miller, 1958; Vermeij,

1973; G. P. Wagner, 1988; Atchley and Hall, 1991; G. P. Wagner, 1995; Cheverud, 1996a; G. P. Wagner, 1996; G. P. Wagner and Altenberg, 1996; Emerson and Hastings, 1998; Bolker, 2000; Eble, 2004; G. P. Wagner and Mezey, 2004; G. P. Wagner et al., 2007; Goswami and Polly, 2010b; Hallgrímsson et al., in press).

Constraints on variation may intuitively seem to prevent long-term evolutionary change, but the constraints imposed by integration and modularity may help coordinate function and developmental associations among traits that would not function well if each sub-trait changed in response to selection in a way that was optimal only for that individual trait and not for the functional or developmental complex. However, as functional and developmental units evolve, integration among traits must also change, lest variation be constrained by obsolete relationships. It has been demonstrated that integration and modularity differ across taxa and, by inference, that patterns change over time (Ackermann and Cheverud, 2000; Badyaev and Foresman, 2000; Marroig and Cheverud, 2001; Ackermann and Cheverud, 2004a; Badyaev and Foresman, 2004; Badyaev et al., 2005; Goswami, 2006a, 2006b; Goswami, 2007a; de Oliveira et al., 2009). These changes fall into two straightforward categories: new functional or developmental interactions link traits into new integration patterns, while fragmentation decouples previous relationships, creating new modularity and allowing old partners to vary independently of each other, thereby promoting “evolvability” (G. P. Wagner and Altenberg, 1996). It has been further hypothesized that fragmentation of older integrative patterns has dominated the history of life, with modularity increasing over evolutionary time as a counter-balance to genetic and developmental integration in increasingly complex systems (Vermeij, 1973; G. P. Wagner and Altenberg, 1996). While these questions can be partially addressed with a phylogenetically-broad sample of data from living taxa, paleontological data have much to offer for understanding evolutionary trends in integration and modularity and their effects on morphological evolution, diversity, and evolutionary rates.

Thus far, the thrust of studies on integration and modularity have focused on establishing the relationships between observed patterns of morphological integration and the possible causes of integration (genetic, developmental, and functional). Many studies of morphological integration have been published, from

studies of insect wings (Klingenberg and Zaklan, 2000; Klingenberg et al., 2001a; Klingenberg, 2004, 2009), to growth series of piranhas (Fink and Zelditch, 1996; Zelditch et al., 2001) and birds (Badyaev and Martin, 2000), to mammalian mandibles and dentition (Atchley et al., 1982; Atchley and Hall, 1991; Cheverud et al., 1991; Atchley, 1993; Cheverud et al., 1997; Badyaev and Foresman, 2000; Klingenberg and Leamy, 2001; Klingenberg et al., 2001b; Leamy et al., 2002; Ehrich et al., 2003; Klingenberg et al., 2003; Badyaev and Foresman, 2004; Cheverud, 2004; Cheverud et al., 2004; Klingenberg et al., 2004; Badyaev et al., 2005; Polly, 2005; Klingenberg, 2009; Zelditch et al., 2009), to the mammalian skull and postcranium (Cheverud, 1982; Cheverud et al., 1983; Cheverud, 1988; Zelditch, 1988; Cheverud, 1989; Zelditch and Carmichael, 1989a, 1989b; Cheverud, 1995, 1996a, 1996b; Leamy et al., 1999; Ackermann and Cheverud, 2000; Lieberman et al., 2000a; Marroig and Cheverud, 2001; Strait, 2001; Hallgrímsson et al., 2002; Ackermann and Cheverud, 2004a, 2004b; Hallgrímsson et al., 2004a, 2004b; Marroig et al., 2004; Zelditch and Moscarella, 2004; Bastir and Rosas, 2005; Young and Hallgrímsson, 2005; Bastir and Rosas, 2006; Goswami, 2006a, 2006b, 2007a, 2007b; de Oliveira et al., 2009; Goswami et al., 2009; Drake and Klingenberg, 2010; Goswami and Polly, 2010b; Bennett and Goswami, in press). In addition, many other studies have addressed the relationship between genetic and phenotypic integration and found some correlation between genetic and morphological integration (Cheverud, 1982; Cheverud et al., 1983; Atchley, 1993; Cheverud, 1996b; Cheverud et al., 1997; Cheverud, 2004; Cheverud et al., 2004), although there is not always a perfect correspondence between genetic modules and functional or developmental modules (Klingenberg and Leamy, 2001; Klingenberg et al., 2001b, 2004). A parallel literature on the relationship between genetic and phenotypic integration exists in the quantitative genetics literature on the evolution of **P** and **G** matrices (matrices of phenotypic and additive genetic variances and covariances, respectively), which is fundamentally about the relationship of phenotypic and genetic integration and how patterns of integration evolve (Roff, 2000; Marroig and Cheverud, 2001; Conner, 2002; Steppan et al., 2002; Polly, 2005; Roff and Mousseau, 2005; Arnold et al., 2008; Polly, 2008; Rice, 2008), as is the literature about evolution along genetic or phenotypic lines of least resistance (Arnold,

1992; Schluter, 1996; Marroig and Cheverud, 2005; McGuigan et al., 2005; Renaud et al., 2006; Zelditch et al., 2006).

Phenotypic integration and modularity are, in principle, composed of both genetic and non-genetic components. Studies on living organisms where heritability data are available are able to partition patterns of phenotypic covariation into its genetic and non-genetic components, whereas studies of paleontological data usually cannot. The uncertainty over whether patterns of integration and modularity are genetic or not need not be of particular concern for the paleontologist because the phenotype is still what interacts with the environment and changes over time, regardless of whether those patterns are entirely genetic or have an environmentally plastic component. Indeed, the concepts of canalization (Waddington, 1942; G. P. Wagner et al., 1997; Rice, 1998; Hallgrímsson et al., 2002) and “robustness” (Nijhout, 2002; de Visser et al., 2003; Hermisson and Wagner, 2004; A. Wagner, 2008) are founded on the idea that phenotypic and developmental integration are not always perturbed by genetic mutations. Furthermore, genetic variance and covariance (the **G** matrix) is by definition a component of phenotypic integration (which is described by the **P** matrix), which makes phenotypic correlations a generally good proxy for genetic correlations (Cheverud, 1988; Roff, 1995). The relationship between phenotypic and genetic components allows for the application of genetic hypotheses to data sets for which only morphological data are available, such as most paleontological data sets.

It is now just over 50 years since the publication of Olson and Miller’s book (1958), and studies of morphological integration and modularity have been extended to a wide range of plants and animals and to a diversity of topics spanning most of the breadth of evolutionary biology, from protein structure to deep time (Pigliucci and Preston, 2004; Schlosser and Wagner, 2004). As the studies and hypotheses concerning integration, modularity, and their evolutionary significance have grown more complex, so too have the methods for identifying and comparing them.

In this paper we review some of the many methods available to study phenotypic integration and modularity, illustrating their similarities and differences with a 3D morphometric dataset of Japanese macaque skulls. We first present background about the study of

morphological correlations, geometric morphometrics (especially Procrustes superimposition), principal components analysis, and the effects of size (allometry), all of which underpin most methods for studying integration and modularity. We then discuss three general methods for studying variances and covariances that are often employed for studying integration and modularity, but which are also used in other contexts: matrix correlation, random skewers, and fluctuating asymmetry. Next we review several approaches for exploring data and discovering patterns of integration and modularity in it, including cluster analysis, Euclidean distance matrix analysis, and graphical modeling. Then we describe approaches for statistically testing whether a new set of data supports existing ideas about patterns, covering partial least squares analysis, integration matrices, theoretical correlation matrix analysis, and RV coefficient analysis. Finally, we consider a method using eigenvalues as an overall index of the degree of modularization in a data set. Our worked example with macaques follows, in which we apply many of the methods we describe to the same data set so their results can be compared. We close by discussing the strengths and weaknesses of each approach, with hopes of providing a starting point for further exploration of this rich topic.

PRELIMINARY CONSIDERATIONS

Covariances and Correlations.—The mathematical key to the quantitative analysis of integration and modularity is correlation among morphological parts. Integration is about the degree of independence among morphological traits and modularity is about whether some subsets of traits are more dependent on one another than they are with traits in other subsets; correlation is an intuitive mathematical measure for independence and dependence. Many measures of correlation have been devised, but all of them express the degree to which two traits change together.

The two most common measures of correlation are covariance and the product-moment correlation (Pearson's R). The covariance is that part of the joint variance of two traits that is associated; the correlation is the same as the covariance for two traits that have been standardized to have variances of one. The units of covariance are the product of the units with which the traits were measured, and its magnitude is related to

the magnitude of the variances. The correlation coefficient is in standardized units and varies between 1.0 (or -1.0) when the two traits covary perfectly and 0.0 when they do not covary at all. The covariance, while less intuitive than the correlation, retains information about the original scale of the two traits that is lost with the correlation coefficient. The covariance between two traits A and B is calculated as:

$$\frac{\sum_{i=1}^n (A_i - \mu_A)(B_i - \mu_B)}{n}, \quad (1)$$

where μ is the mean of its respective trait and n is the number of specimens in the sample (for the covariance of a sample, $n - 1$ is used). The product-moment correlation coefficient is simply the covariance standardized by the joint variances of the two traits:

$$\frac{\sum_{i=1}^n (A_i - \mu_A)(B_i - \mu_B)}{n \cdot s_A s_B}, \quad (2)$$

where s_A and s_B are the standard deviations of the two traits.

When many traits are studied, as is usually the case with modularity, the correlations among those traits are organized in a covariance or correlation matrix for analysis (Figure 1, A-B). For m traits (or variables), such a matrix has m rows and m columns, where the diagonal cells contain the variances of each of the m traits and the off diagonal elements contain their covariances (for a correlation matrix, the diagonal elements are scaled to equal variance, 1.0, and their off-diagonal elements are correlations). Covariance and correlation matrices are symmetrical, meaning that if the columns and rows are transposed the matrix remains unchanged, and they are non-negative definite, meaning roughly that the eigenvalues of the matrix are all positive or zero (see below). If the data are distributed linearly and multivariately normal with respect to one another, then the covariance matrix can be thought of as a description of the dispersion of data in the trait space. More importantly for the present paper, the covariance and correlation matrices describe the integration among the traits, "packages" of which can be identified using the methods described below. In the example in Figure 1 there is only one "package"

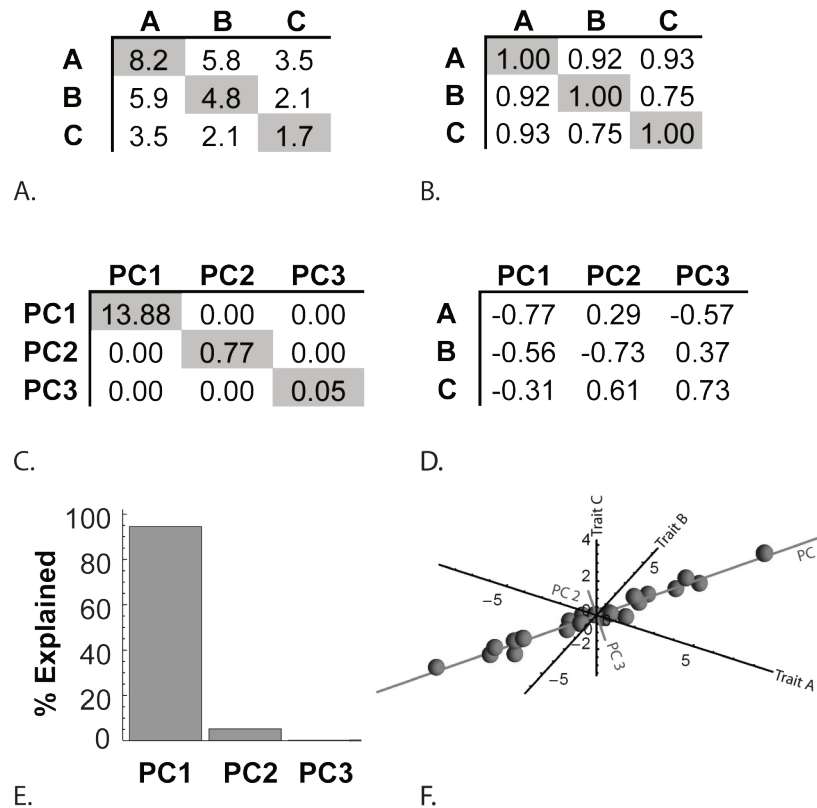


FIGURE 1.—**A.** Covariance matrix for three traits *A*, *B*, and *C*. The diagonal elements are the variances, and the off-diagonal elements are the covariances. **B.** Correlation matrix for the same three traits. Off-diagonal elements are the product-moment correlations among the traits. **C.** Eigenvalues for the principal components (eigenvectors) of the covariance matrix (diagonal elements). Notice that the covariances among the three principal components are all zero (off-diagonal elements) and that the sum of the eigenvalues is equal to the sum of the variances in **A**. **D.** Eigenvectors of the covariance matrix. These numbers can be thought of as loadings of each trait on each principal component, or as angle (in radians) that each principal component must be turned to be aligned with the original variable axes. **E.** The percentage of the sum the original variances explained by each principal component. **F.** The spatial relationships between data, trait axes, and principal components. Traits *A*, *B*, and *C* are correlated, so lie in a linear cloud in the 3D space formed by their trait axes. The principal components are the major axes through those data, lying at angles to the original trait axes described by the eigenvectors in **D** and having variances along each principal component axis described by the eigenvalues in **C**.

of correlation, or one “module”.

Most geometric morphometric analyses are based on covariances of Procrustes superimposed landmarks rather than correlations because the variables of interest are the *x*, *y*, and *z* coordinates of landmarks and the scaling between coordinates is important to retain lest shape be distorted (Bookstein, 1991; Rohlf, 1993; Zelditch et al., 2004). The correlation coefficient does not retain that scaling because the variances, and hence the scales, of both variables are standardized to 1.0.

Conversely, most analyses of integration and modularity use correlations because the issue of interest is not whether the absolute scale of covariation is large or small, but whether two components of the phenotype covary proportionally more or less than other components (Olson and Miller, 1958; Cheverud, 1995; G. P. Wagner et al., 2007).

For studying modularity, we prefer to use correlations among Procrustes superimposed landmarks rather than covariances because we are more interested in the

packaging of integration among parts of the phenotype than in the geometric representation of the covariation of the shape. Because landmark coordinates are multivariate, either two- or three-dimensional, choices must be made among several methods for calculating the correlation. Some authors treat each landmark coordinate separately, which has the analytical advantage of not losing any shape information, but has the interpretative disadvantage that different coordinates of the same landmark may be linked to different modules, a situation that is plausible in some biological situations but implausible in others. We prefer to treat landmarks as complete units (unless the particular question being addressed demands that landmarks be treated as individual coordinates, which is not the case in any of our examples).

Two common, closely-related coefficients exist for calculating correlations among multidimensional variables: the canonical correlation (Hotelling, 1936) and the congruence coefficient (Burt, 1948), both of which are extensions of Pearson's R to two multivariate blocks of data. The canonical correlation coefficient is the correlation between the scores of two sets of variables on their respective major axes, the congruence coefficient is the sum of the covariances between the two multidimensional variables over all the specimens in the dataset, divided by the pooled variance (Abdi, 2007). We use the congruence coefficient as applied to 3D landmarks in our analyses, except where noted, and we refer to it with the designation R (referring to it as the correlation between landmarks when more than one level of correlation is being discussed). We calculated R as follows:

$$R_{ij} = \frac{\sum (l_i - \mu_i) \cdot (l_j - \mu_j)}{\sqrt{\sum l_i^2 \cdot \sum l_j^2}}, \quad (3)$$

where n is the number of specimens in the sample, l_i and l_j are 3D vectors of the x-y-z coordinates of landmarks i and j , μ_i and μ_j are the centroids of those landmarks, and the lower term is the pooled sample variances of the two landmarks. Note that this equation uses the dot product in the numerator. This is the same landmark correlation that we have used elsewhere (Goswami, 2006a, 2006b, 2007a; Goswami and Polly, 2010a).

Klingenberg (2008a) criticized this coefficient, asserting that it did not detect correlations except in

cases where landmarks covary in the same anatomical direction. While Klingenberg's suggestion has some merit, in fact the congruence coefficient does detect covariances in opposite directions (but R becomes negative, as it should). The only situation where Klingenberg's criticism pertains is when variation at both landmarks is highly linear (i.e., the variation is primarily in a single direction, not scattered in the x , y , and z directions, a pattern we have not yet observed in our studies of within-species variation) and when the axis of variation of one landmark occurs in a direction that is at or near 90° to the variation in the second landmark. In that limited situation, a high correlation will be underestimated by the congruence coefficient. In other situations, the loss of information from the congruence coefficient is small, and the gain from representing a biological landmark as a single unit of information outweighs that loss. Every method for analyzing multivariate landmarks has its mathematically strong and weak points, and these must be considered when choosing a method or comparing results of different analyses. Regardless of our choice of correlation coefficient, the methods we present below could be used with any method of measuring correlation between variables unless otherwise noted.

Procrustes Superimposition. — An important step in most analyses of geometric shape is the registration of landmarks to a common orientation and scale using Procrustes superimposition (Rohlf, 1990; Rohlf and Slice, 1990; Dryden and Mardia, 1998; Zelditch et al., 2004). In generalized Procrustes analysis (GPA), which is described in detail by Webster and Sheets (this volume), the landmark configurations for each specimen in the analysis are standardized to unit size, translated to the origin of the coordinate system, and rotated with respect to the sample mean until the sum of distances among the specimens is minimized. In some variants on this standard procedure the original scaling of the landmarks is retained so that the resulting Procrustes coordinates contain information about both shape and size variation (often called "form" as a distinction from "shape"), although this may induce size-related correlations among neighboring landmarks if alternative treatments of size variation are not applied. Two practical consequences of Procrustes superimposition are relevant to the study of modularity and integration.

The first important consequence of Procrustes

superimposition is that degrees of freedom are lost because of the Procrustes superimposition. For any set of landmarks, the number of coordinate variables is $2k$ for two-dimensional landmarks (2D) or $3k$ for three-dimensional (3D) ones, where k is the number of landmarks. Procrustes reduces the mathematical and statistical degrees of freedom by 4 for 2D data (one for scale, one for translation along the x -axis, one for translation along the y -axis, and one for rotation in a plane) and by 7 for 3D data (one for scale, one each for translation in the x , y , and z -axes, and three for rotation in a three dimensional space). The reduction in degrees of freedom causes a covariance or correlation matrix of the landmarks to be “singular” (or, more technically, “non-positive definite”), meaning that the rows and columns are not independent of one another and the inverse of the matrix cannot be calculated. Many quantitative methods of analysis rely on the inverse of a correlation or covariance matrix, including the method of graphical analysis that is reviewed below (Magwene, 2001, 2008). Such methods may be impossible to apply to Procrustes superimposed landmarks unless an alternative algorithm is found (often one involving a “pseudoinverse” matrix).

The other consequence of Procrustes superimposition is that the variance at any particular landmark is a product of both biological variation and the least-squares solution that minimizes distances among all landmarks (Chapman, 1990; Rohlf and Slice, 1990; Richtsmeier and Lele, 1993; Lele and Richtsmeier, 2001). The use of interlandmark correlations for the study of integration and modularity largely circumvents the issue because the correlation is a measure of the relative position of one landmark to others, not its absolute position. Methods such as EDMA and the analysis of interlandmark distances explicitly address the uncertainties of superimposition (see below).

Principal Components Analysis.—Principal components analysis (PCA) is a method normally used to reduce the dimensionality of multivariate data (Jolicoeur and Mosimann, 1960; Jolliffe, 2002), but in the study of modularity and integration PCA is used to reorganize data into axes each of which describe intercorrelations among specific groups of variables. Principal components (PCs) are a set of axes, the first of which is the major axis of the multivariate cloud of data. Subsequent axes are at right angles to prior

components, meaning that variation along the subsequent components is the uncorrelated residual variation around the prior axes. Each PC thus describes correlated variation among several of the original variables that is uncorrelated with variation on other principal components (this is mathematically true, but see Mitteroecker et al., 2004 for ways in which it may not be biologically true.)

The mathematical descriptors of principal components are eigenvalues and eigenvectors. The eigenvalues are the amount of the original variance explained by each PC (Figure 1C). The sum of the eigenvalues is equal to the sum of their corresponding covariance matrix so that collectively they account for all the original variance in the data. If the eigenvalues are written in diagonal matrix form, the off-diagonal elements are all zero because there is no correlation (or covariance) among the PCs. Thus, the eigenvalues can be thought of as a transformed covariance matrix in which the new variables have variance but no covariance. The eigenvectors describe the transformation from the new PC axes to the original variable axes (Figure 1D). The eigenvector for each PC has a loading for each variable. Large absolute values indicate a high correlation, either positive or negative, between the variable and the PC. Variation along each eigenvector accounts for a certain proportion of the original variance, the largest portion of it on the first PC (Figure 1E). The more covariation there is among the original variables, the more variance is explained by the first few principal components. In this example, the correlations among the original variables are high, so the data points form a tight, almost linear scatter in the variable space (Figure 1F). Thus the strong correlations between X on the one hand and Y and Z on the other ($R = 0.92$ and 0.93 respectively) result in 94% of the variation lying along the first principal component. The variation in this example data set is highly integrated such that it can be described almost in its entirety by a single PC variable, which can be thought of as a single module.

The numbers making up the eigenvectors are actually the angles, in radians, between the PC axis and the original variable axis. In the example in Figure 1 the angle between PC 1 and the x -axis is -0.77 radians, or -43.9 degrees. Visual inspection of Figure 1D shows that the two lines are, indeed, lying at nearly a 45 degree angle from one another. Collectively the eigenvector matrix is a rotation matrix that rotates data rigidly from

the original variables to the PC axes or back from the PCs to the original variables. Because of the lack of correlation among axes of the PC space and the ease with which positions in that space can be transformed back to the original variable space, the PC space makes a convenient tool for simulating or otherwise analyzing multivariate phenotypic evolution (Polly, 2004). For the purposes of modularity and integration, the eigenvector of each PC is a simple measure of modularity while the dispersion of the eigenvalues is an easily comparable measure of integration (see below).

Allometry.—Shape often varies predictably with size as an animal grows and, thus, allometry, the size-related component of shape variation, is important to consider in any analysis of shape, including those concerning integration and modularity. In particular, because size often affects an entire shape, whether it is the entire the cranium or entire organism, it can potentially create the appearance of complete integration and mask modularity. While Procrustes superimposition partitions size out of an analysis, it does not remove the component of shape that is correlated with size. Readers should consider whether allometry is important to their interpretations because any source of covariance, be it genetic, functional, or size-related, will contribute to the appearance of integration and modularity. Allometric covariation may genuinely be a product of modularity and integration or it may create the appearance of integration where there is none; interpretations could be misguided either by retaining size-related variation or by removing it. In our example study below, we expect that growth-related changes in cranial correlations may be channelled through integrated modules, so we do not remove shape variation by regressing against size across age cohorts, which would remove evidence of those modules that channel growth-related changes. Instead, we analyzed each age cohorts separately so that size-related variation contributes relatively little to correlations within a data set.

Should it be desirable to remove allometry, this is best done by replacing raw Procrustes coordinates with the residuals of a regression analysis of the Procrustes coordinates against centroid size, or some other body size proxy, prior to calculating a covariance or correlation matrix (Klingenberg, 2009).

Matrix correlation analysis.—Matrix correlation analysis is a widely used method for assessing the level

of correspondence between two matrices (Sneath and Sokal, 1973). It is simply the Pearson's product moment correlation, described above, extended to two correlation matrices. Matrix correlation analysis is often used in studies of morphological integration for comparing correlation matrix structure between different sets of specimens, whether different populations, ages, or species, or different types of correlations, such as genetic, phenotypic, and environmental (Cheverud, 1995). Mantel's test is often used to compare the structure of correlation matrices (Mantel, 1967). Mantel's test is a permutation test that assesses if the similarity between two matrices is significantly greater than would be expected from two random matrices of equal size. It accomplishes this by holding one matrix constant, while randomly permuting the rows and related columns of the second matrix and recalculating the matrix correlation between the original and the permuted matrix (only one set of off diagonal elements are compared since the two sets of off-diagonals are identical and the diagonal elements always equal 1.0 in a correlation matrix). Repeating this exercise several times, often 10,000 times, creates a null distribution against which to assess the original matrix correlation. Generally, significance is established if the observed matrix correlation is greater than 95% of the randomly generated matrix correlations (Cheverud et al., 1989).

It should also be noted that, while matrix correlations theoretically range from +1 to -1, sampling error makes it essentially impossible to achieve perfect correlations between matrices by reducing the true matrix correlation. To correct for the effect of sampling error, Cheverud (1996b) suggests adjusting observed matrix correlations by dividing them by maximum achievable matrix correlation, calculated as the square root of the product of the repeatabilities of each matrix.

Random skewers analysis.—Random skewers analysis is an alternative matrix comparison method which compares the responses of two covariance matrices to randomly-generated selection vectors as a measure of similarity (Pielou, 1984; Cheverud, 1989; 1996b; Marroig and Cheverud, 2001). The selection vectors are generated by randomly drawing elements from a distribution of -1 to 1 and then normalizing the vector to a total length of 1. The vector correlation between the response vectors from the two matrices is then calculated. The analysis is repeated, perhaps 1000 times, and the vector correlations are then averaged

to provide a single measure of similarity between the two matrices of interest. As with matrix correlation analysis, the significance of the mean vector correlation produced by random skewers analysis is assessed by comparison with a distribution of vector correlations generated by random vectors (Cheverud, 1996b). As a guide, for the 61-element vector, used here, a vector correlation of 0.31 would be considered significantly greater than zero at the $p < 0.01$ level, in comparison to vector correlations from a random distribution of 1000 vectors. Originally used in ecological analyses (Pielou, 1984), this method has been used successfully in several studies of phenotypic covariance/variance matrices (Cheverud et al., 1983; Cheverud, 1996b; Marroig and Cheverud, 2001). One can also use this approach to assess covariance matrix repeatability, or the likelihood that you will get the same matrix from a different sample of the same population, by repeatedly resampling the original dataset and generating a mean vector correlation between the original and resampled datasets. Covariance matrix repeatability of each matrix can then be used to adjust the observed vector correlations for the effects of sampling error, as described above for matrix correlations.

Fluctuating Asymmetry.—Fluctuating asymmetry (FA) is the random difference in what are otherwise bilaterally or radially symmetrical traits (Van Valen, 1962; Palmer and Strobeck, 1986). Because the left and right sides of an organism are presumably under the same genetic and developmental controls, traits that are randomly asymmetrical result from environmental rather than genetic perturbations. Such perturbations could be external to the organism, such as the effects of handedness in humans or the effects of gravity and physical pressure from the position in the uterus in Eutherian mammals, or they could be internal, such as random binding problems in signaling proteins during development. Regardless, the phenotypic differences associated with these perturbations are unquestionably non-genetic (sometimes described as being the result of “developmental noise”, “ecophenotypic variation”, or “plasticity”), and can be considered such even when the data normally needed to partition genetic and non-genetic variation are unavailable. When a suite of traits is influenced by the same developmental pathways, then asymmetric environmental perturbations will result in a correlated change in those traits on one side relative to another. Thus, randomly covarying groups

of asymmetric traits can be used to identify modules or patterns of integration that are created by shared developmental processes.

Klingenberg et al. (2002; Klingenberg, 2003) described a method to identify modules based on the shared covariance among asymmetrical landmarks. To do this, bilateral data are collected from a series of individuals where each individual is measured several times to control for error. First the landmarks on one side are reflected across the axis of body symmetry and Procrustes aligned with the landmarks of the opposite side. A two-factor ANOVA is then performed to partition the variation among individuals, between sides, and between replicates (Palmer and Strobeck, 1986). Landmark covariance matrices are then constructed for each partition and a principal components analysis performed on the among-individual and between-sides matrices to explore the overall pattern of integration and the pattern of correlation in the asymmetrical variation (Klingenberg et al., 2002). Correlated landmarks identified from the interaction of the among-individuals and between-sides data partitions may belong to developmentally integrated modules because they are the ones that shift in a coordinated fashion in response to random effects between the left and right sides. Similarity between the among-individuals component and individual sides (FA) component indicates that the symmetric component of phenotypic variation was influenced by direct interactions among developmental pathways (Drake and Klingenberg, 2010).

Even though we have described fluctuating asymmetry separately, the symmetric and asymmetric components of variation can be analyzed with most of the methods described below so long as the phenotypic variation is first partitioned into these components.

EXPLORATORY METHODS FOR IDENTIFYING MODULES AND INTEGRATION

In the following sections we first review some of the existing approaches for identifying patterns of modularity and integration and then discuss some methods for comparing and testing these patterns. The division of methods into exploratory and evaluative is imperfect, because some methods, such as Partial Least Squares analysis, could be listed under either topic. Our review is not complete and, in some cases, the different

approaches that we review overlap with one another. One notable method that we have not described here is Common Principal Components analysis (see Phillips and Arnold, 1999 and references therein), as our previous studies have shown this to be unstable at sample sizes below 35 (Polly, 2005).

Most of the existing approaches identify patterns of integration and modularity in a phenotypic correlation (or covariance) matrix, extracting that pattern using one or more methods to identify structures that are more correlated with one another than with other structures. The differences between approaches are mostly in which data are used to construct the phenotypic correlation matrix, how those data are standardized, which methods and statistical criteria are used to identify modules or integration patterns, and whether the approach aims to identify patterns in a single sample (e.g., a population or species) or to identify shared patterns in many samples.

Cluster analysis.—Cluster analysis is a versatile exploratory method for assigning landmarks (or other items) into groups based on a chosen measure of similarity, and it is a component of many of the methods used to study integration and modularity (Legendre and Legendre, 1998). Most cluster methods arrange items into hierarchically arranged sets, with smaller clusters grouped into successively larger ones. Users of cluster analysis are required to define both a distance metric and a linkage algorithm. Many distance metrics are available, but Euclidean distance is the most commonly used one in morphometrics because it describes the shortest distance between two items that have been described with continuous variables. In morphometrics, Euclidean distances can be calculated among objects (such as specimens or group means) using their superimposed landmark coordinates or their principal components scores (Q mode analysis), or Euclidean distances can be calculated among traits using the loadings of the traits on the principal components of a correlation matrix or directly from the correlation matrix by subtracting the absolute value of each correlation from 1.0 (R mode analysis) (Legendre and Legendre, 1998).

The most common linkage algorithms in morphometrics are unweighted pair-group average (UPGMA), Nearest Neighbour (or single linkage), and Ward's method (Ward Jr, 1963; Sneath and Sokal, 1973).

UPGMA joins groups based on the mean distance between all elements in each group, while Nearest Neighbour uses only the smallest distance between elements in each group. Ward's method differs from both as it optimizes for within-group variance, rather than raw distance, in creating clusters, specifically using the error sum of squares around the group mean as a measure of fit and then finding the pattern that minimizes the within-group variance of each cluster (Ward Jr, 1963). Choice of linkage algorithm varies among users, but Ward's method has been used widely in identifying potential modules (Cheverud, 1982), as described further below.

Zelditch et al. (2009) demonstrated some shortcomings of the use of cluster analyses in identifying integration when the relationships among landmarks are not hierarchical. They specifically cite a low cophenetic correlation for the dendrogram they produced in a study of the mouse mandible (the cophenetic correlation is the correlation of distances on the tree with the original morphometric distances [Sneath and Sokal, 1973]). Thus, cluster analyses may not be well suited to all datasets, and the agglomeration coefficient (Kaufman and Rousseeuw, 1990) can be used to estimate if one's data are sufficiently hierarchically ordered to support use of a cluster analysis. UPGMA trees, especially when they are constructed using an ultrametric tree algorithm, are known for distorting distances, whereas other methods may perform much better (Polly, 2001). It is worth noting that cluster analyses will always find clusters, even when there is little structure to a dataset, so it is important to test for the significance of within-cluster trait correlations before describing them as integrated.

Euclidean Distance Matrix Analysis.—Procrustes superimposition distributes variance among the landmarks in whatever manner best minimizes the sum of squared distances between the shapes, a process that may reduce the perceived variance of highly variable landmarks or increase the perceived variance of stable ones, making it difficult to localize which parts of a shape are more variable than others (Rohlf, 1990; Richtsmeier et al., 2002). Some approaches to modularity and integration are, therefore, based on the matrix of pairwise interlandmark distances instead of the superimposed landmark coordinates.

The Euclidean Distance Matrix Analysis (EDMA) method of shape analysis is based on the invariant prop-

erties of the interlandmark distance matrix, often known as the “form matrix”, under rotation and translation to study shape variation (Lele and Richtsmeier, 1991; Richtsmeier and Lele, 1993; Cole and Richtsmeier, 1998; Lele, 1999; Lele and Richtsmeier, 2001). EDMA avoids superimposition, which by necessity uses a non-biologically homologous reference point (the geometric mean of the landmark coordinates) to align landmarks, and instead compares the ratio of two form matrices to get at shape differences. If two forms differ purely in size, then the ratio of all interlandmark distances will be the same, but if the forms differ in shape, some of the ratios will be significantly different than others. The coordinates of the mean shape (or form) of a set of objects are reconstructed from their average form matrix. EDMA has been used in various contexts to study morphological integration and modular change (Hallgrímsson et al., 2004a; Boughner et al., 2008; Bruner, 2008; Bruner and Ripani, 2008).

EDMA is associated with shape spaces whose topography is more irregular than the shape spaces associated with Procrustes superimposed landmarks (Rohlf, 2000, 2003), which makes the estimation of mean landmark configurations, evolutionary trajectories, or developmental sequences through shape space challenging for EDMA. Likewise, the effects of Procrustes superimposition on the redistribution of landmark variance make it challenging to localize particular landmarks that are more variable than others (Lele and Richtsmeier, 1991; Richtsmeier and Lele, 1993; Cole and Richtsmeier, 1998; Lele, 1999; Lele and Richtsmeier, 2001; Richtsmeier et al., 2002), which makes the study of anything other than variation in an entire shape difficult with Procrustes methods. Both methods are equally good for studying correlations among landmarks, with the caveat that size variation can be treated differently in the two methods, which makes them both useful approaches for studying modularity and integration. It should be noted, however, that methods based on Procrustes superimposition usually identify modules made up of a subset of landmarks, whereas methods related to EDMA usually identify modules made up of connections between landmarks. The difference between the two is subtle, but it is possible that contradictory results from the methods are due to them having measured different things.

Cheverud (1982) introduced a widely cited method for studying modularity and integration that is

conceptually and methodologically related to EDMA in that it is based on matrices of interlandmark distances rather than on Procrustes superimposed landmark coordinates. In Cheverud’s method, the correlation matrix of the interlandmark distances was constructed and the principal components (eigenvalues and eigenvectors) of that matrix were obtained. Cheverud then used the loadings of the traits on the first 20 principal components axes (those components with positive eigenvalues) as coordinates of the traits in a 20-dimensional variable space. An inter-trait distance matrix was constructed and Ward’s clustering method, described above, was applied to identify modular associations among the traits. Individual clusters were matched to cranial regions of interest by the mode number of traits associated with that region (e.g., if a cluster had 6 out of 8 traits belonging to the facial region, it was identified as a facial cluster). Cheverud’s method has either been adopted by or has influenced many subsequent studies (e.g., Goswami, 2006a, 2006b; Zelditch et al., 2009).

Graphical modeling.—In a series of elegant papers, Magwene (2001, 2008, 2009) applied graphical modeling techniques to the study of integration and modularity. Graphical modeling is a special branch of statistics that seeks to represent conditional independence between variables using graphs in which the variables are represented by vertices and high partial correlations between pairs of variables by lines connecting their vertices (Whittaker, 1990; Lauritzen, 1996). When applied to the problem of morphological integration, traits are represented by graph vertices and the correlations by lines connecting them. Modules appear as subsets of connected graph vertices, visually capturing the conceptual module of what a module should be. Zelditch et al. (2008) used a closely related method for studying connections among modules, which they refer to as a reticulated network. These methods differ from cluster analysis in that they do not impose a strictly hierarchical pattern on relationships among landmarks or modules, allowing multiple interactions to be detected.

In practical terms, the key to graphical modeling is the inverse of the trait correlation matrix $\Omega = \mathbf{P}^{-1}$, where \mathbf{P} is the correlation matrix (which is identical to \mathbf{P} , the phenotypic correlation matrix mentioned above) and Ω is its matrix inversion (Magwene, 2001). The diagonal elements of Ω equal $1/(1-R^2)$, where R^2 is the multiple correlation coefficient between that trait and

all the others. When Ω is rescaled so that the diagonals equal 1.0, then the off-diagonal elements are the negative partial correlation coefficients between pairs of variables. Partial correlations that are near zero are discarded using one of several statistical criteria, including the edge exclusion deviance (a criterion derived from information theory that uses χ^2 distribution for calculating statistical probability) and edge strength (effectively a cut-off procedure based on the natural logs of the partial correlations), leaving the higher positive or negative correlations as edges on the graph.

Mitteroecker and Bookstein (2007) criticized the graphical modeling method, arguing that it works only when there are exactly two traits in a module because otherwise the correlations among traits within a module are cancelled out by their conditional relationship to other traits within the module. Magwene (2009) rebutted this criticism, demonstrating that it recovered modular structures consisting of many traits. The main issue, as with all the methods described here, is whether sample sizes are large enough to accurately measure correlations and partial correlations (Mitteroecker and Bookstein, 2009). In practice, the graphical modeling method cannot usually be applied to correlation matrices derived from Procrustes superimposed landmarks because the matrix is singular, having lost $2k - 4$ or $3k - 7$ degrees of freedom from scaling, translation and rotation for 2D and 3D landmarks respectively, where k is the number of landmarks (Dryden and Mardia, 1998). Singular matrices cannot be reliably inverted, though generalized inverses are available, as are other possible solutions that would allow the graphical modeling method to be applied to landmark data, including “bending” the matrix so that it is no longer singular. As shown below, the issue of singularity does not arise if links among modules identified from integration matrices are being studied (Zelditch et al., 2009)

CONFIRMATORY METHODS FOR TESTING MODULES AND INTEGRATION

Two-block Partial Least Squares and Singular Warps.—Two-block partial least squares (PLS) and singular warps are methods that can be used to compare integration and modularity in different taxa. At its most fundamental level, morphological integration and modularity concerns covariances within an individual,

population or species because it is at that level where shared developmental, genetic, and functional factors will result in between-trait correlation or independence. But the question of whether patterns of correlation and independence change from one group to another is also fundamental to the notion of whether modularity and integration direct and constrain the path of evolution, or whether they themselves are evolvable (Klingenberg, 2005; Draghi and Wagner, 2007; Jones et al., 2007; Hansen and Houle, 2008; Polly, 2008). Comparisons among correlation matrices and phenotypic datasets are necessary to understand how modularity and integration change.

PLS is a method for extracting the eigenvectors and eigenvalues from two blocks of covarying data (Rohlf and Corti, 2000) and has been extended to more than two blocks (Streissguth et al., 1993; Bookstein et al., 2003). When applied to geometric shape and thin-plate splines, Bookstein et al. (2003) refer to the PLS method as Singular Warps Analysis. If the two sets of variables are combined and a correlation matrix R constructed from them, then

$$R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}, \quad (4)$$

where R_{11} is the block of correlations among the first set of variables, R_{22} is the block of correlations among the second set of variables, and R_{21} and R_{12} are the two cross-blocks of correlations between the two sets of variables (where $R_{12} = R_{21}'$). PLS finds the principal components of covariation between the datasets by decomposing R_{21} into eigenvalues and into two sets of eigenvectors, one for each data set, using singular value decomposition to produce $F_1 D F_2'$, where F_1 and F_2 are the matrices of eigenvectors for the two datasets and D is the matrix of singular values, which are the square-roots of the eigenvalues (Rohlf and Corti, 2000). As with principal components, the first axis explains most of the covariation between the datasets, but note that whether it explains a significant part of the total variation in the combined datasets is a different matter. Also note that the first PLS axis may explain more of the variation in one data set than the other. If the data sets are the same landmarks applied to two different species (or other group), then PLS can be used to study the modularity and integration shared by those two groups.

Integration Matrices.—Monteiro and colleagues

introduced the concept of “integration matrices” to analyze the integration among modules that had been identified *a priori* (Monteiro et al., 2005). The core of Monteiro’s method was to look at patterns of integration among pre-defined components of the morphology that consist of multiple landmarks. The data used by Monteiro et al. (2005) were Procrustes superimposed landmarks of a complex structure consisting of several subunits (they studied the rodent mandible, which was divided by Atchley and Hall (1991) into condylar, coronoid, angular, and alveolar morphogenetic components based on *a priori* knowledge of developmental tissue interactions). For each subunit, Monteiro et al. (2005) calculated a matrix of pair-wise distances between the landmarks of that subunit on each pair of individuals in the sample as a measure of the distribution of variation of that unit in the population. They then calculated pair-wise matrix correlations between the distance matrices, with the expectation that integrated subunits will have higher correlations than loosely integrated ones; the matrix of those pair-wise correlations among subunits they called the integration matrix. Monteiro et al. (2005) used clustering to illustrate the patterns of integration among subunits and used partial least squares (PLS, described above) to determine whether the pattern of integration found in one taxon or hierarchical level of taxa was similar to integration in other taxa or at other levels. This method has been fruitfully combined with the fluctuating asymmetry method to study sources of integration in living and fossil organisms (Webster and Zelditch, 2008; Zelditch et al., 2009).

RV coefficient analysis. — Recently, Klingenberg (2009) introduced a robust new method based on Escoufier’s RV coefficient (Escoufier, 1973) to test whether hypothesized modules fit observed covariance matrices better than a distribution of randomly-generated partitions of equivalent size and number. The RV coefficient is analogous to a correlation coefficient, but instead of measuring the association between two individual variables, it measures the association between two sets of variables. The calculation of the RV coefficient involves essentially dividing the covariance between the two sets of variables by the variance of each set, as in the calculation of a correlation (the RV is closely related to the congruence coefficient that we use to measure the correlation between 2D and 3D landmarks, except that the RV coefficient is for symmetric matrices

and ranges from 0 to 1, whereas the congruence coefficient is for non-symmetric matrices and ranges from -1 to 1 (Abdi, 2007)). The calculation of the RV coefficient for two sets of variables, as described in detail by Klingenberg (2009), is as follows. If one defines a covariance matrix, S ,

$$S = \begin{bmatrix} S_1 & S_{12} \\ S_{21} & S_2 \end{bmatrix}, \quad (5)$$

where S_1 and S_2 are the separate covariance matrices for the two sets of variables being tested, and S_{12} is the covariance matrix between those two sets, the RV coefficient is:

$$RV = \frac{\text{trace}(S_{12}S_{21})}{\sqrt{\text{trace}(S_1S_1)\text{trace}(S_2S_2)}}. \quad (6)$$

The numerator of the above equation reduces to the sum of the squared covariances between the two sets of variables, while the denominator is the square root of the product of squared total variance of each set. Klingenberg (2009) further expands the RV coefficient approach to allow comparisons of multiple sets of variables, basically by averaging all pairwise RV coefficients among sets. The RV coefficient ranges between zero and one, with zero representing a case of perfect modularity, in which there is no covariance between the sets.

In order to test the significance of the RV coefficient or multi-set RV coefficient, Klingenberg (2009) devised an impressive permutation test that compares the observed RV coefficient for the hypothesized modules to those of different randomly-generated but equally-sized partitions of the same landmarks. He further recommends limiting random partitions to spatially contiguous sets of traits, with the argument that biological modules, particularly those related to developmental units, are more likely to be spatially contiguous. Although this does not necessarily apply to functional modules, or even to all developmental modules, using entirely random partitions will inevitably produce sets of traits that do not make any anatomical sense. This will effectively bias the significance test towards finding the RV coefficient of the hypothesized and hopefully anatomically realistic modules to be significantly lower than the null distribution. However, one should base the decision on whether to limit partitions to spatially contiguous sets on the attributes of their particular system. Even holding

the number and size of partitions equal and limiting partitions to spatially contiguous sets still generates a large number of possibilities for large or complex datasets, but a large set of random partitions (10,000 is recommended by Klingenberg, 2009) should be sufficient for determining if the observed RV coefficient is significantly lower, reflecting low covariance between the hypothesized modules, than expected from the null distribution. RV coefficient analysis as described above can be conducted with the MorphoJ software package (Klingenberg, 2008b).

Theoretical matrix analysis.—While the matrix comparison methods described above are mainly used for assessing the similarity of two samples, they are also useful for testing hypothesized patterns of integration (as indeed are most of the methods described here). One common method for doing so is by construction of a theoretical connectivity matrix with the hypothesized pattern of integration and then comparing it to an observed correlation matrix with matrix correlation analysis and Mantel's test for significance. A connectivity matrix is a square, symmetric matrix of 1s and 0s that has the same layout as a correlation matrix. A 1 is entered in cells of the connectivity matrix where the two variables belong to the same module and a 0 is entered where they belong to different modules. Several examples of this approach exist, generally testing for correspondence between functional units or developmental structures, such as mesenchymal condensations, and phenotypic integration (Cheverud, 1995; Marroig and Cheverud, 2004; Goswami, 2006a). Simply, one creates a theoretical connectivity matrix by assigning correlations of 1 between landmarks with a hypothesized relationship and 0 between those without a hypothesized relationship. As Cheverud (1995) notes, the connectivity matrix represents the ideal correlation matrix, but in reality, it is improbable that functionally or developmentally integrated traits would show a perfect correlation of 1. Similarly, many traits that are not biologically integrated may have non-zero correlations. Thus, perfect correspondence between the theoretical and observed correlation matrices should not be expected.

For datasets with multiple possible sets of relationships, such as different modules, one can test each one separately by including only one set in a connectivity matrix, or one can pool all of the possible

sets into a total integration matrix. Separate analyses would be useful in cases where some, but not all, of the hypothesized relationships are supported in the observed correlation matrix. However, these analyses do not provide a measure of best fit for selecting among the different models in cases of multiple comparisons, but would simply show which, if any, of the theoretical patterns of integration are consistent with the observed correlation matrix (Cheverud 1995).

INDICES OF OVERALL MODULARITY AND INTEGRATION

Eigenvalue dispersion.—As discussed above in the section on principal components analysis, large datasets can be decomposed into a few axes that describe the majority of variation in the dataset. These axes, the principal components, are combinations of the original variables in the dataset that reflect the covariation among variables, in effect reflecting which variables are strongly integrated with each other. The eigenvalues are a measure of how much of the variance in a dataset is explained by each principal component. If many of the original variables in a dataset covary strongly with each other, as one would expect in a strongly integrated unit, then most of the variance in a dataset will be contained within one or a few principal components. If only one or a few principal components have very large eigenvalues, and the remaining axes have very low eigenvalues, eigenvalue variance will be high. However, if the structure of a dataset is strongly modular, with several different groups of strongly covarying traits, then variance will be distributed more evenly across a number of principal components, and eigenvalue variance will be relatively low. Thus, the dispersion of eigenvalues provides a simple measure for comparison of the relative integration or modularity of the structures described in a matrix.

The standard metric used in analysis of morphological integration is scaled eigenvalue variance. Because the measure of interest for comparisons across datasets is the difference among eigenvectors, not the total variance of a sample, all eigenvalues are first scaled by dividing by their mean eigenvalue, and then their variance is calculated. A recently suggested alternative is to compare eigenvalues from a correlation matrix, rather than a covariance matrix, because correlation matrices are scaled to equal variance and thus

retain information only about intertrait relationships, the key factor of interest in studies of morphological integration (Pavlicev et al., 2009). They further suggest that the relative standard deviation of eigenvalues from correlation matrices best represents the structure of a dataset and provide a metric that is independent of trait numbers, which is particularly useful in comparisons of disparate datasets.

$$\text{Var}(\lambda) = \frac{\sum_{i=1}^N (\lambda_i - 1)^2}{N} \quad (7)$$

$$\text{SD}_{\text{REL}}(\lambda) = \frac{\sqrt{\text{Var}(\lambda)}}{\sqrt{N-1}} \quad (8)$$

Note that the mean eigenvalue in Equation 7 is replaced by 1 because the sum of the eigenvalues for a correlation matrix is the same as its trace, which is equal to the number of traits. Thus the mean eigenvalue is essentially the number of traits divided by the number of traits, or 1.

While eigenvalue variance and relative standard deviation both provide a simple and easily interpretable measure of overall integration for comparison across different datasets, the main limitation of eigenvalue dispersion is that it provides relatively little information about how matrices differ or the actual patterns of biological integration.

MODULARITY AND INTEGRATION IN MACAQUE SKULLS: A WORKED EXAMPLE

Materials and Methods.—To provide examples of some of these analyses and to compare their results, we performed covariance and correlation matrix comparisons, analysis of eigenvalue dispersion, analysis of interlandmark distances, and tested three models of integration using a dataset of a dataset of 61 3D landmarks (Table 1, Figure 2) gathered from 181 cranial specimens of Japanese macaques, *Macaca fuscata fuscata*, housed in the Primate Research Institute at Inuyama, Japan. The specimens were divided into four age groups based on tooth eruption: 1) deciduous dentition (42 specimens); 2) adult M1 erupted (42 specimens); 3) adult M2 erupted (48 specimens); 4) all adult dentition erupted (49 specimens). The adult specimens were subdivided into female (24 specimens) and male (25 specimens) datasets, but lack of gender data

prevented the pre-adult specimens from being similarly divided. A maximum of two specimens in each age sample were subjected to a mirroring algorithm to fill in missing data in symmetrical landmarks. These four datasets were then subjected to generalized Procrustes analysis and a total of nine outliers (1 deciduous, 3 M1, 1 M2, and 4 adult) were identified (with a Principal Components Analysis and by measuring Procrustes distance from the mean shape) and removed, prior to being used in analyses.

Correlation matrices generated from 3D landmarks with the congruence coefficient of correlation were used in the following analyses: cluster analyses of landmarks, integration matrices, graphical modeling of integration matrices, matrix correlation analysis, random skewers analysis, eigenvalue dispersion analysis, and theoretical matrix correlation analysis. All analyses were conducted in Mathematica 7.0 (Wolfram Research Inc., Champaign, Illinois). Fluctuating asymmetry and RV coefficient analysis were performed using MorphoJ (Klingenberg, 2008b), in which, unlike in the other analyses conducted here, landmark coordinates are treated separately in generating covariance matrices, but not in permutation tests. Analysis of correlations in interlandmark distances was performed in Mathematica 7.0. Mathematica notebooks with routines used here are available for download at: <http://mypage.iu.edu/~pdpolly/Software.html>. Additional software for various analyses of morphological integration and modularity (e.g. MorphoJ, IMP, Sage, Mace) can be obtained from the SUNY Stony Brook Morphometrics Site: <http://life.bio.sunysb.edu/morph/index.html>.

For analyses in which we tested specific models of modularity (integration matrices, graphical modeling, RV coefficient analysis, and theoretical matrix correlation analysis), three models were used (Table 1). The first is a simple division into two modules, corresponding to the facial region and the neurocranial region, as described in Drake and Klingenberg (2010). The second model, proposed by Cheverud (1995) decomposes the cranium into six modules (oral, nasal, orbit, vault, zygomatic, and base) based on functional and developmental relationships and has been used in multiple analyses of primate crania, including macaques (Cheverud, 1982; Cheverud, 1995). Because Cheverud's modules are based on interlandmark distances, the individual landmarks were assigned to the module in which the majority of their respective

Table 1.— List of 61 landmarks used in analyses of *Macaca fuscata fuscata*, 27 of which are bilaterally paired (*). Landmark associations for each of the three models (N/F, neurocranial-facial 2-module model; Cheverud 6-module model; and Goswami 6-module model) of cranial modularity tested in various analyses are also listed.

No	Description	N/F	Cheverud	Goswami
1	Premaxilla anterior midline suture	1	1	1
2	Premaxilla - Maxilla lateral suture*	1	1	1
3	Premaxilla - Maxilla ventral suture	1	1	1
4	Canine - lateral extreme*	1	1	1
5	Canine - mesial extreme*	1	1	1
6	Palatine - Maxilla - ventral suture	1	1	2
7	Maxilla Palatine lateral suture*	1	1	2
8	Anterior P1*	1	1	2
9	Nasal - anterior extreme	1	2	1
10	Nasals - Frontal suture	1	3	3
11	Nasal - Premaxilla - Nasal opening suture*	1	2	1
12	Jugal-Maxilla (Orbit crest) suture*	1	3	3
13	Lacrima - Frontal - Maxilla suture*	1	3	3
14	Ethmoid - Lacrima - Frontal suture*	1	3	3
15	Maxilla jugal posterior interior*	1	4	3
16	Sphenoid jugal frontal interior*	1	4	3
17	Jugal-Maxilla (base of zygoma) suture*	1	4	2
18	Jugal-Frontal (postorb bar) suture (crest)*	1	3	3
19	Jugal-Squamosal dorsal suture*	1	4	4
20	Jugal-Squamosal ventral suture*	1	4	4
21	Parietal Frontal suture	2	5	5
22	Parietal - Squamosal - Frontal suture*	2	5	5
23	Parietal Squamosal Occipital*	2	5	5
24	Jugal frontal alisphenoid*	2	4	5
25	Frontal sphenoid squamosal*	2	4	5
26	Pterygoid tip lateral*	2	6	4
27	Pterygoid tip medial*	2	6	4
28	Presphenoid - Palatine - Alisphenoid*	2	6	4
29	Basisphenoid - Presphenoid - Alisphenoid*	2	6	4
30	Basioccipital-Basisphenoid-Bulla suture*	2	6	6
31	Bulla anterior medial extreme*	2	6	6
32	Bulla posterior lateral extreme*	2	6	6
33	Parietal - Occipital suture	2	5	4
34	Occipital condyle lateral extreme*	2	6	6

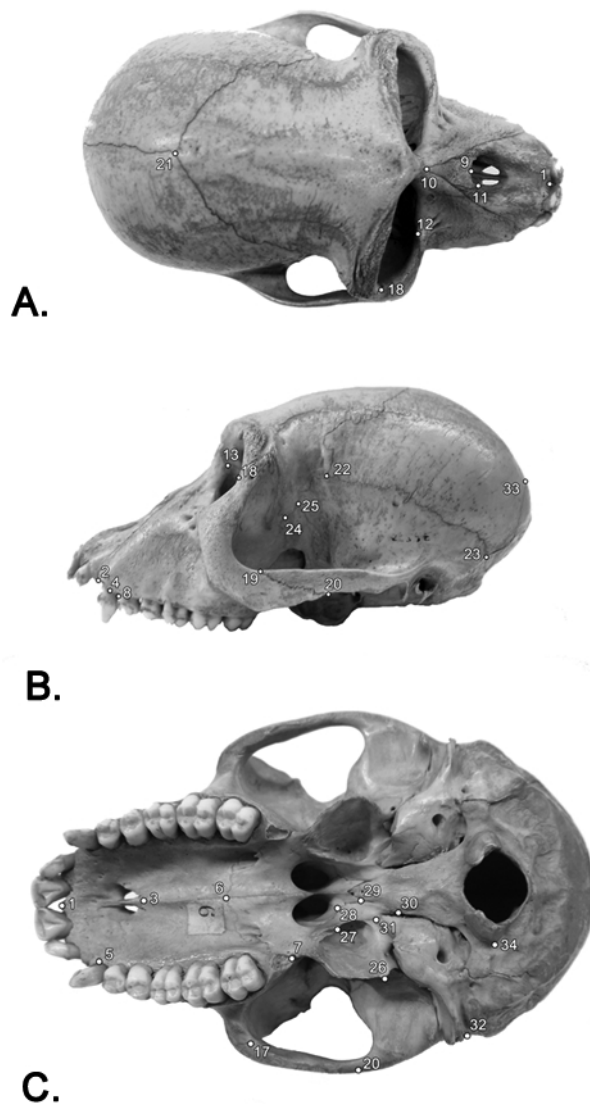


Figure 2.— The landmarks used in this study. A. Dorsal view. B. Lateral view. C. Ventral view. Landmarks are numbered as in Table 1. (note that the 27 bilaterally paired landmarks are only illustrated on one side).

distances were involved. The last model is a similar six-module model derived from exploratory cluster analyses of several species of marsupials, carnivorans, and primates by Goswami (2006a) that identified modules corresponding to the anterior oral-nasal region, the molar-palatal region, the orbit, the zygomatic-pterygoid region, the vault, and the basicranium.

Sample size issues.—As is well appreciated, large sample sizes are ideal for morphometric analyses, particularly those where covariance matrices have to be estimated. Unfortunately, large sample sizes can be difficult to obtain in studies involving paleontological specimens or rare extant taxa. For this reason, we performed a subsampling analysis (Sanders, 1968; Raup, 1975) of the deciduous-only macaque dataset to assess the effects of small sample sizes on analyses of correlation structure. From the original 41-specimen dataset, 1 to 31 specimens were removed, with 100 repetitions of each step conducted, for a total of 3100 runs. Subsampled datasets were subjected to generalized Procrustes analysis following removal of specimens, and then a correlation matrix was generated for each new dataset as described above. The subsampled correlation matrices were then compared to the original correlation matrix with matrix correlation analysis and Mantel's test (Steppan, 1997; Polly, 2005).

Another issue with broad comparative analyses, or analyses including paleontological data, is the lack of specimen information, such as gender. For taxa with sexual dimorphism, this can introduce an additional and important complication. Realistically, gender data are a luxury that paleontologists will rarely enjoy, and thus it important to try to estimate the effects of mix-gender samples on analyses of covariance and correlation matrices. To address this issue, we compare the results of gender-pooled, adult male only and adult female only analyses and discuss the effect that pooling male and female specimens may have on analyses.

Subsampling analysis.—The effects of small sample size on the estimation of landmark correlations was assessed through subsampling of the deciduous dataset, where the number of specimens was incrementally and randomly decreased from 41 to 10 and the resulting correlation matrices compared to the original with matrix correlation. The correspondence with the original dataset stayed high, average matrix correlation greater than 0.90, with as few as 14 specimens, similar to the results of Polly (2005). Even at low sample sizes, matrix correlations between the subsampled and original datasets were mostly higher than matrix correlations between different age groups, whose range is shown as the shaded region in Figure 3. Even with only 10 specimens, the matrix correlations between subsampled samples and the original were

also mostly higher than the mean unadjusted matrix correlation between eight different species of *Macaca*, reported as 0.80 by de Oliveira et al. (2009), suggesting that even very small sample sizes, while far from ideal, are sufficient for estimating differences in landmark correlation structure between age cohorts or species.

Matrix correlation analysis and random skewers analysis.—Both matrix correlation analysis and random skewers analysis showed a decrease in similarity of correlation or covariance structure, respectively, with increasing difference in age between groups (Table 2). The pattern of landmark correlation of the juveniles with deciduous dentitions, for example, was most similar to the subadults with M1 ($R=0.81$) and least similar to pattern of correlation in adults ($R=0.73$, 0.70 and 0.67 for the pooled adult, adult female and adult male samples respectively). However, even the cohorts most different in age were highly significantly similar in the pattern of landmark correlation ($p < 0.01$). The landmark correlation structure of the pooled adult sample was almost more strongly similar to younger cohorts than were the individual adult male and female samples. Neither the adult female nor male samples were consistently more similar to the younger cohorts than the other gender.

It should also be noted that covariance matrices of the asymmetric and symmetric components of shape were significantly correlated in each age group, though correlations increased with age (Table 2), suggesting directions of shape variation across individuals correspond to patterns of variation driven by developmental interactions.

Cluster analysis of landmarks.—Modules were identified from the correlation among Procrustes superimposed landmarks from the adult female dataset in two ways. First a dissimilarity matrix was calculated directly from the correlation matrix as $1-\text{abs}(R)$, where $\text{abs}(R)$ is the absolute value of R , the correlation coefficient. Ward's clustering was done on the resulting matrices and the top six clusters were retained (Figure 4A-F). Second a pair-wise distance matrix of variables was calculated by measuring the distance of landmarks from one another in the variable space defined by the first 21 eigenvectors of the correlation matrix, which always describe more than 99% of the total variation in our datasets (Figure 4G-L). Note that the regions rep-

resented by different colours in Figures 4 and 5 reflect traditional divisions of the skull and do not correspond perfectly to either of the six-module models. The six clusters are not listed in any particular order.

Neither method returned modules that were identical to those found in previous studies; both approaches returned clusters combining elements of most or all of modules found in previous studies. The cluster analysis of the dissimilarity matrix produced a more modular structure than the distance matrix of variables, in that only two of the clusters mix more than two previously-identified modules, whereas all of the distance matrix-based clusters mix at least three modules. The four least-inclusive clusters (those that involve the fewest landmark) generated from the dissimilarity matrix generally correspond to a neurocranial-facial division. Cluster 2 is essentially a nasal-orbit cluster, while Cluster 5 is mainly an anterior oral cluster (along with a rogue landmark from the cranial vault). Clusters 3 and 6 resemble the zygomatic-pterygoid module identified by Goswami (2006a), but the similarities are generally quite weak. Little other structure is apparent.

Interlandmark Distance Analysis.—Modules were also identified using the correlations among interlandmark distances, which include size as well as

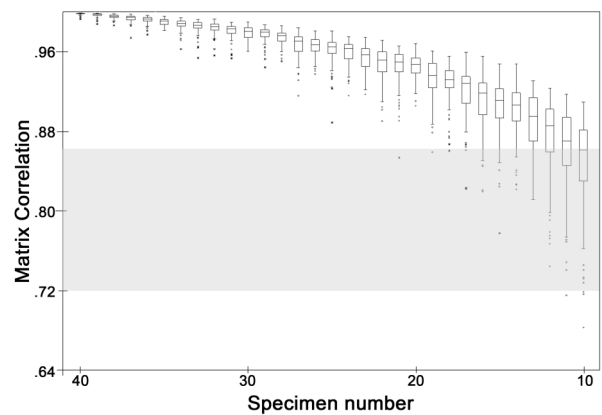


FIGURE 3.—Results of subsampling test of the effects of sample size on the estimation of landmark correlation structure. Box plots represent the range of matrix correlations with the original dataset for 100 resampled matrices at the respective sample size. Shaded region indicates the range of matrix correlations between the different age groups of *Macaca fuscata* studied here.

TABLE 2.— Matrix correlation coefficients between six correlation matrices for 61 landmarks. Results from random skewers analysis (upper triangle, grey) and matrix correlation analysis (lower triangle, white). Diagonal elements are the matrix correlations between the asymmetric and symmetric covariance matrices for each age group. All results are significant at the $p < 0.01$ level, as determined by a permutation test (10,000 runs).

	Deciduous	M1	M2	Adult all	Adult female	Adult male
Deciduous	0.287	0.81	0.77	0.73	0.70	0.67
M1	0.81	0.394	0.83	0.83	0.76	0.80
M2	0.77	0.83	0.399	0.85	0.78	0.81
Adult all	0.72	0.83	0.86	0.493	0.92	0.94
Adult female	0.69	0.76	0.79	0.91	0.400	0.77
Adult male	0.66	0.80	0.82	0.94	0.78	0.345

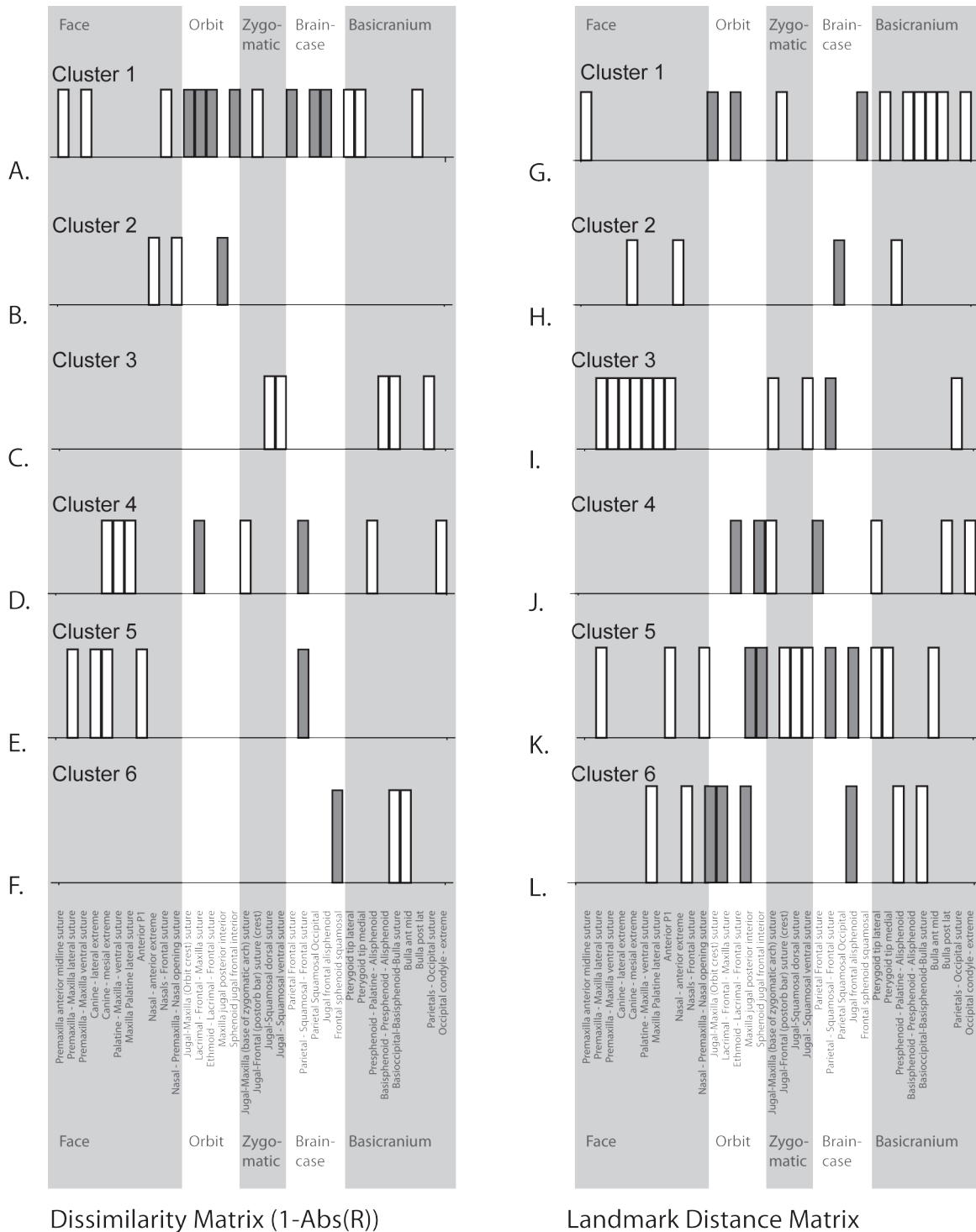
shape, again from the adult female dataset. To do this, all pair-wise distances among landmarks were calculated, a correlation matrix was calculated, eigenvectors of that matrix were calculated, and clusters were found with Ward's method from the inter-variable distances in the first 21 dimensions of the variable space.

Six clusters were investigated as potential modules (Figure 5) based on the break in consistency of within-module correlations compared to within-module correlations at the next higher hierarchical level (the criterion used by Cheverud (1982)). Because each interlandmark distance involves two landmarks and each landmark is involved in many pair-wise distances, the involvement of each landmark in a module was summarized by calculating how many module distances it contributed to in the module and dividing by the total number of distances that landmark contributed to in the entire dataset. These modules are interpreted as follows. Module 2 is the face and the anterior part of the braincase. Module 3 involves the overall length or flexion of the skull because the high-scoring landmarks are arranged from the bridge of the nose back to the occipital condyle. Modules 1, 4, 5 and 6 are all braincase modules (even though the landmark in Module 4 is colored blue with the basicranium, it is the junction of the parietals and occipital and so is related functionally to the braincase). Thus, the interlandmark distances have identified what could be considered a two module system: face-basicranium, braincase, and interaction between the two. The integration of the face and basicranium is consistent with the observation that the cranial base and facial region exert great influence on each other during growth in primates (Lieberman et

al., 2000a, 2000b; Ackermann and Cheverud, 2004a). The segregation of the interlandmark distances involving the braincase across four different clusters perhaps reflects the lower integration of the cranial vault in primates compared to other mammals, which has also been noted previously (Cheverud, 1996a; Ackermann and Cheverud, 2004a; Goswami, 2006a)

There is little similarity between these results and the results of the landmark-based dissimilarity and distance matrices. The closest similarity is that there is a facial-neurocranial division, though none of the landmark-based analyses recovered a facial-basicranial interaction to the exclusion of the braincase.

Integration matrices.—We used integration matrices (Monteiro et al., 2005) to look at the pattern of integration among the six modules identified by Goswami (2006a), detailed in Table 1, in each of the four ontogenetic cohorts of *Macaca*. For each data set, landmarks were superimposed then divided into their respective modules. For each module, a pair-wise inter-specimen distance matrix was calculated as the Procrustes distance of the landmarks in that module. An integration matrix built from the pair-wise matrix correlations between those distance matrices and clusters were found from the integration matrices using Ward's method (Figure 6A-D). The integration patterns changed substantially from the deciduous cohort to the subadult and adults, similar to the findings from matrix comparisons (discussed above) and eigenvalue dispersion (discussed below), although the neurocranial modules (vault and basicranium) maintain a close relationship throughout. Note that procedures were



also described by Monteiro et al. (2005) for statistically testing whether the correlations within modules are stronger than those between by partitioning the putative modules, but we have not applied those tests here.

Graphical modeling.—Because the relationship among modules does not need to be strictly hierarchical, as represented in a cluster diagram, Magwene's (2001, 2008) graphical modeling method was also applied to the integration matrices for comparison to the Ward's dendrograms (Figure 6E-H). Integration was highest, in that there was more interconnection among modules, in the deciduous cohort, as suggested below by the analysis of eigenvalue dispersion, and it declined through successive cohorts.

RV coefficient analysis.—RV coefficients were generated for both the symmetric and asymmetric components of variation for the three models described above (Table 1). None of the models was significantly correlated with the asymmetric covariance matrices, but each was significant ($p < 0.05$) for some, but not all, of the symmetric matrices (Table 3). The neurocranial-facial model was significant for the M1, M2, pooled adult, and adult male datasets. The Cheverud model was significant for the deciduous, M1, M2, and pooled adult datasets, and the similar Goswami model was significant for the M2, pooled adult, and adult male datasets. While all three had relatively similar significance values for the analyses of the adult dataset, the Cheverud model was much more significant ($p \leq 0.001$) for all of the subadult datasets.

It should be noted that the RV coefficient was much higher for the neurocranial-facial division than for the six module models, suggesting that there is greater correlation between these two large modules than, on average, among the six modules in the other two models. The higher significance of the two-module model may not necessarily indicate a better fit than the six-module models because different numbers of parameters are involved, and graphical modelling or structural equation analysis using the Akaike information criterion could be used to further test which model better fits the data.

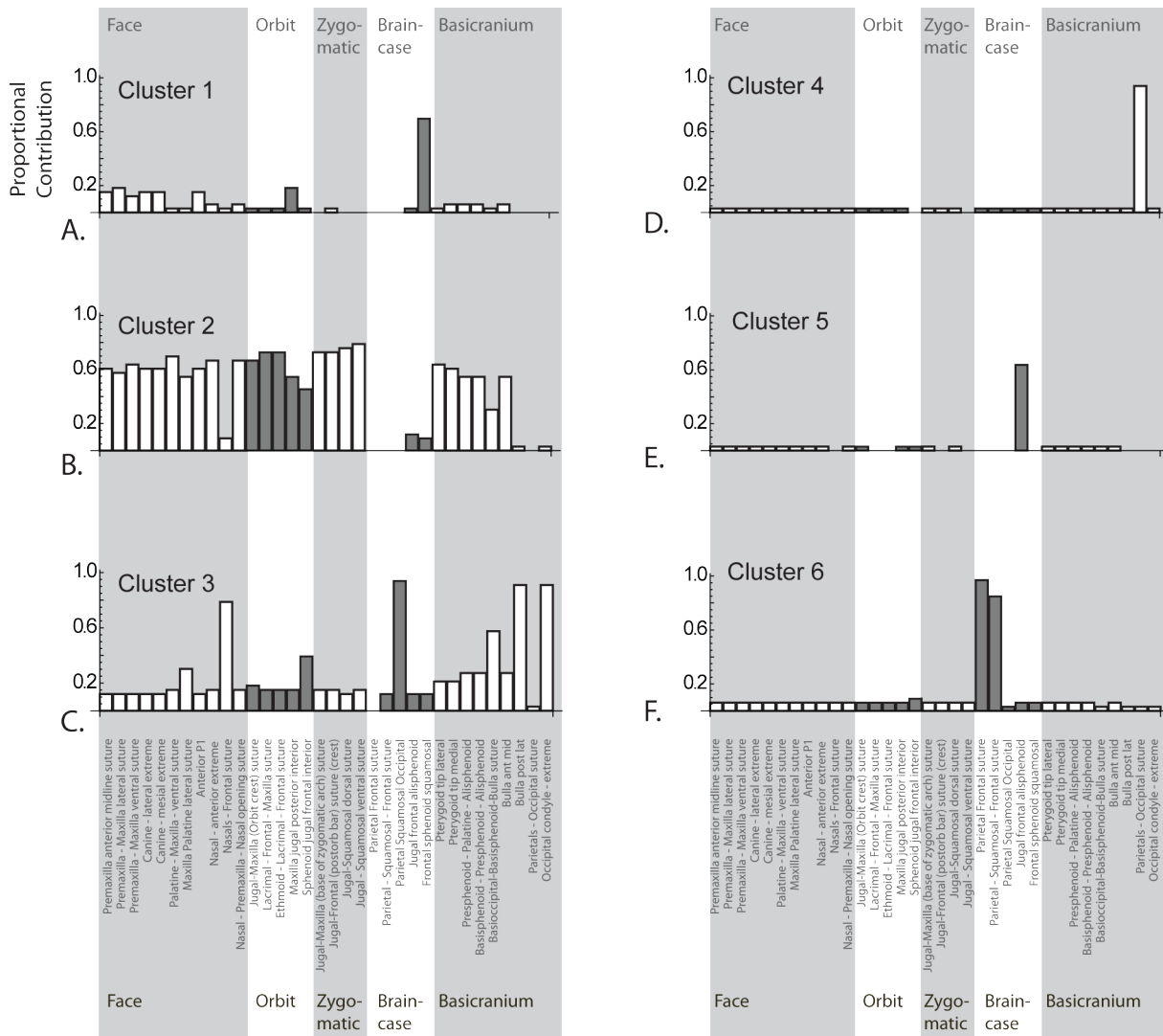
Theoretical matrix correlation analysis.—As with RV coefficient analysis, theoretical matrix correlation analysis was applied to three pre-existing models

of cranial integration (listed in Table 1). Results of the theoretical matrix correlation analyses show that nearly all of the three models are significantly correlated with the observed matrices as determined by Mantel's test (10,000 runs, $p < 0.001$) (Table 4). Matrix correlations were strongest with Cheverud's six-module model, and lowest with the two-module neurocranial-facial division model.

Eigenvalue dispersion.—The variance of eigenvalues from each cohort's correlation matrix was quantified for each of the four age groups and for the samples of adult males and females (Table 5). The eigenvalue variance was highest in the youngest age cohort, those with only deciduous dentition, suggesting that group had the highest level of integration, but there was little difference among the others, suggesting that the overall level of integration, if not the precise pattern of integration, remains relatively constant across ontogeny in this species.

Comparison of Results.—As mentioned in the introduction to this chapter, there are a wide array of methods used to examine morphological integration and modularity. Here, we have attempted to describe and compare some of the most common approaches, and we find that, though there are some differences, many methods return relatively similar results.

The greatest differences among the results produced by parallel methods were among the three different types of cluster analysis used to identify modules. Although some biological structure was apparent in the analysis of the landmark-based dissimilarity matrix and interlandmark distance analysis, neither of the sets of modules identified satisfactorily matched each other or the modules expected based on previous functional and developmental studies. The clusters based on interlandmark distances had the additional complication of containing landmarks from different cranial regions, making identification of modular structure problematic. Because both the RV coefficient and theoretical matrix analyses found statistically significant correspondence between existing models of modularity and our *Macaca* data, it seems unlikely to us that the cluster analyses are reliably identifying the biological associations among traits. Empirical studies of modularity that use clustering should also include careful statistical assessment of whether the modules that are identified



Analysis of interlandmark distances (distances in variable space)

FIGURE 5.— Results of cluster analysis of interlandmark distances, showing the landmarks included in each cluster as vertical bars that are color-coded by topographic region of the skull (landmark names and color-codes are at the bottom of the figure). A.-F. Six clusters identified by the distances in a variable space constructed from the correlation matrix of interlandmark distances. Because each interlandmark distance involves two landmarks and because each landmark contributes to many interlandmark distances, the contribution of each landmark is shown by the height of its bar (the number of distances the landmark contributes to in that cluster divided by the total number of distances it contributes to in the dataset).

TABLE 3.—RV coefficients and p-values for the three pre-existing models of cranial integration, analysed for symmetric and asymmetric components. Bold entries are significant at the $p < 0.05$ level.

		Neuro-Face	<i>P</i>	Cheverud	<i>P</i>	Goswami	<i>P</i>
Deciduous	Sym	0.75	0.163	0.371	0.001	0.481	0.13
	Asy	0.303	0.577	0.203	0.633	0.173	0.114
M1 erupted	Sym	0.556	0.017	0.31	0.001	0.359	0.119
	Asy	0.388	0.124	0.302	0.979	0.259	0.701
M2 erupted	Sym	0.664	0.02	0.349	0.001	0.402	0.014
	Asy	0.32	0.056	0.189	0.657	0.202	0.135
Adult	Sym	0.514	0.014	0.303	0.016	0.313	0.02
	Asy	0.424	0.409	0.249	0.85	0.235	0.604
Adult Female	Sym	0.504	0.057	0.351	0.299	0.368	0.491
	Asy	0.537	0.441	0.305	0.745	0.318	0.87
Adult Male	Sym	0.543	0.002	0.412	0.438	0.369	0.033
	Asy	0.615	0.702	0.402	0.81	0.391	0.678

are well-supported by the data and whether they better explain patterns of phenotypic correlation than do existing models of modularity.

Our results suggest that graphical modeling as a promising approach for studying trait associations because it does not impose strictly hierarchical relationships and may be a better model for the relationships among modules than a dendrogram. Other network approaches have the same strength (e.g., Zelditch et al., 2009). As noted above, graphical modeling is problematic when used with Procrustes-transformed landmarks because it cannot tolerate singular covariance matrices. However, there are a number of methods to circumvent this issue. We demonstrated on such approach here, using integration matrices instead of covariance or correlation matrices. Other possible solutions include matrix bending (Phillips and Arnold, 1999), using a pseudoinverse of the covariance matrix, or using a subset of PC scores instead of landmarks (and then translating the results back to the original landmarks for interpretation).

The results of the analyses other than the clustering were generally congruent with one another.

Graphical modeling of the integration matrices and eigenvalue dispersion both identified the deciduous dentition cohort as the most integrated, with integration decreasing in older age groups. The increase in modularity with age identified by these two methods is illuminated by the concurrent increase in the correlation between the asymmetric and symmetric components of variation with age, perhaps suggesting that canalization, represented by decreasing symmetric variation, and developmental stability, measured as a decrease in fluctuating asymmetry, are coordinated and together produce a more modular adult phenotype rather than in earlier stages of development.

The matrix comparison methods, random skewers analysis and matrix correlation analysis returned very parallel results: the similarity in correlation structure decreased with increasing age difference. Matrix correlation analysis directly compares the elements of the two matrices in question, while random skewers measures the similarity of their response to random vectors. Matrices that are more similar in elemental composition will also have more similar responses to vectors, so it is not surprising that the two methods

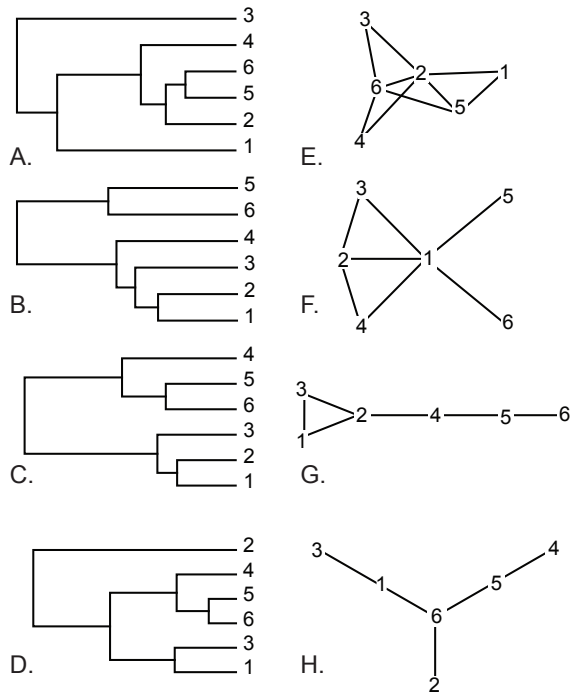


FIGURE 6.— Analysis of integration matrices among the six empirical modules identified by Goswami (2006a) using Ward's clustering (A-D) and graphical modeling (E-H). A,E. Cohort with deciduous dentition. B,F. Cohort with one erupted adult molar. C,G. Cohort with two erupted adult molars. D,H. Adult cohort (both sexes).

produce similar results, but Cheverud (1989) argued that matrix correlation analysis is inappropriate for analysis of covariance matrices because the scale is not equal across all of its elements and recommended using random skewers analysis instead.

The two methods used to test *a priori* models of modularity, theoretical matrix correlation analysis and RV coefficient analysis, returned similar results despite being based on matrices generated with different treatments of landmark coordinates in calculating the measure of correlation. Statistical significance was generally lower in RV coefficient analysis, but all three models were significant in the pooled adult sample for both analyses.

TABLE 4.—Results of theoretical matrix correlation analysis for each of three *a priori* models of cranial modularity: a simple two-module division between the neurocranium and face, a theoretical modular pattern proposed by Cheverud for testing with primate cranial data and an empirical modular pattern identified by Goswami from marsupials, carnivores and primates. (* indicates $p < 0.01$, ** indicates $p < 0.001$).

	Neuro-Face	Cheverud	Goswami
Deciduous	0.224**	0.524**	0.414**
M1 erupted	0.296**	0.564**	0.426**
M2 erupted	0.204**	0.573**	0.409**
Adults all	0.191**	0.585**	0.463**
Adult females	0.152*	0.556**	0.413**
Adult males	0.202**	0.546**	0.439**

TABLE 5.—Relative standard deviation of eigenvalues ($SD_{REL}(\lambda)$) of the correlation matrices of the four age cohorts and for the adult female and male samples.

Age Cohort	$SD_{REL}(\lambda)$
Deciduous	0.255
M1 erupted	0.217
M2 erupted	0.242
Adults all	0.219
Adult females	0.225
Adult males	0.245

FURTHER CONSIDERATIONS

Overall, we found that the exploratory approaches (cluster analyses of correlations among landmarks and interlandmark distances) were less informative and less consistent with one another than were the results of model testing or comparative approaches. The one limitation we found with the model testing approaches was that the models being tested were often similar enough that they were not statistically distinguishable, and for this reason we expect that in many cases several models may be significantly correlated with an observed matrix. This lack of differentiation among models may reflect biological reality,

however, as many different and overlapping layers of developmental interactions may be influencing trait correlations. Hallgrímsson et al. (in press) have termed this the “palimpsest model” of cranial integration, in analogy to reused scrolls on which the remnants of previous communications obfuscate a clear reading of the text. Márquez (2008) recently described a method that simultaneously tests multiple models of modularity and also combines them and considers complex interactions across models. Large numbers of independent and combined models can be ranked by goodness of fit, and consensus models can be constructed from the best supported models. This promising approach has not yet been extended to three-dimensional data, but the ability to test and compare hundreds of models appears to be a useful intermediate between blind exploratory approaches, such as cluster analysis, and the limited comparisons among individual models possible with existing model-testing methods.

As noted above, sample sizes are almost always limited in analyses involving paleontological specimens. Aside from the general advice that the more specimens, the better, we hesitate to provide specific recommendations for what constitutes an adequate sample size, because the minimum acceptable sample size depends on the size of the effect that is being measured relative to the variation in the population. For comparative analyses, we can suggest that running a subsampling analysis on the largest possible sample, as we did here, is a useful way to assess whether the question of interest can be addressed with smaller samples.

Another sampling aspect that we touched on in passing was the question of sexual dimorphism. The degree to which lack of gender data is a problem again depends heavily on the data set, but differences between males and females can be substantial. In the dataset used here, matrix correlations between male and female Japanese macaques are similar to the differences between age groups. The pooled adult sample had lower integration and stronger correspondence to the other age groups, which were also pooled due to lack of gender information, than the individual male and female samples. Greater variance might be expected to return higher measures of integration (Hallgrímsson et al., in press), but that does not appear to be case here. Our pooled adult sample also generally showed higher correlations with the hypothesized modules and lower

and more significant RV coefficients than the individual gender samples. However, both of individual gender samples also returned significant results for all of the theoretical correlation matrix analyses, and the adult male, but not the adult female, results were consistent with those of the pooled adult dataset for two of the three RV coefficient analyses. Thus, with the exception of RV coefficient analysis, the differences between males and females in this sexually dimorphic species appear to have relatively small effect on the interpretation of results. There is also the practical issue that gender data is nearly impossible to obtain for extinct taxa, and, for this reason, we feel that pooling genders should be considered a necessary evil that is unlikely to distort the results of large-scale analyses.

Lastly, we recommend using the correlation matrix and its respective analyses in studies of morphological integration and modularity, because, as discussed above, the attribute of particular interest is the relationships among traits, rather than the relative variance of those traits (Pavlicev et al., 2009).

An increasing number of studies of the evolution of phenotypic modularity and morphological integration have included broader and broader comparisons among the phylogenetic diversity of living organisms, but there have not yet been many studies that incorporate paleontological data to directly measure evolutionary changes. It is hoped that this volume will inspire more paleontologists to consider these issues, first approached by Olson and Miller (1958) in what remains as one of the few studies of morphological integration in fossil taxa.

ACKNOWLEDGMENTS

The authors thank John Alroy and Gene Hunt for inviting them to participate in this short course, for making helpful comments that improved the paper, and for their patience with editing. Mark Webster and an anonymous reviewer made thoughtful observations on this manuscript. Data were gathered at the Primate Research Institute in Inuyama, Japan, with funding provided by the Japanese Society for the Promotion of Science HOPE grant to AG. We thank M. Takai for his role in encouraging and facilitating the research trip to PRI.

REFERENCES

- ABDI, H. 2007. The RV coefficient and the congruence coefficient. *In* N. Salkind (ed.), *Encyclopedia of Measurement and Statistics*. Sage, Thousand Oaks, CA.
- 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 components analysis. *American Journal of Physical Anthropology*, 111:489-501.
- ACKERMANN, R. R., AND J. M. CHEVERUD. 2004a. Morphological integration in primate evolution., p. 302-319. *In* M. Pigliucci and K. Preston (eds.), *Phenotypic Integration*. Oxford University Press, Oxford.
- ACKERMANN, R. R., AND J. M. CHEVERUD. 2004b. Detecting genetic drift versus selection in human evolution. *Proceedings of the National Academy of Sciences USA*, 101: 17946-17951.
- ARNOLD, S. 1992. Constraints on phenotypic evolution. *American Naturalist*, 140:S85.
- ARNOLD, S., R. BURGER, P. HOHENLOHE, B. AJIE, AND A. JONES. 2008. Understanding the evolution and stability of the G-matrix. *Evolution*, 62:2451-2461.
- ATCHLEY, W. R. 1993. Genetic and developmental aspects of variability in the mammalian mandible, p. 207-247. *In* J. Hanken and B. K. Hall (eds.), *The Skull*. Volume 1. Development. University of Chicago Press, Chicago.
- ATCHLEY, W. R., AND B. K. HALL. 1991. A model for development and evolution of complex morphological structures. *Biological Reviews*, 66:101-157.
- ATCHLEY, W. R., J. J. RUTLEDGE, AND D. E. COWLEY. 1982. A multivariate statistical analysis of direct and correlated response to selection in the rat. *Evolution*, 36:677-698.
- BADYAEV, A. V., AND K. R. FORESMAN. 2000. Extreme environmental change and evolution: stress-induced morphological variation is strongly concordant with patterns of evolutionary divergence in shrew mandibles. *Proceedings of the Royal Society of London Biological Sciences, Series B*, 267:371-377.
- 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*). *Journal of Evolutionary Biology*, 13:290-301.
- BADYAEV, A. V., AND K. R. FORESMAN. 2004. Evolution of morphological integration. I. Functional units channel stress-induced variation in shrew mandibles. *American Naturalist*, 163:868-879.
- BADYAEV, A. V., K. R. FORESMAN, AND R. L. YOUNG. 2005. Evolution of morphological integration: developmental accommodation of stress-induced variation. *American Naturalist*, 166:382-395.
- BASTIR, M., AND A. ROSAS. 2005. The hierarchical nature of morphological integration and modularity in the human posterior face. *American Journal of Physical Anthropology*, 128:26-34.
- BASTIR, M., AND A. ROSAS. 2006. Correlated variation between the lateral basicranium and the face: a geometric morphometric study in different human groups. *Archives of Oral Biology*, 51:814-824.
- BENNETT, C. V., AND A. GOSWAMI. in press. Does reproductive strategy drive limb integration in marsupials and monotremes? *Mammalian Biology*.
- BOLKER, J. A. 2000. Modularity in development and why it matters to evo-devo. *American Zoologist*, 40:770-776.
- BOOKSTEIN, F. L. 1991. *Morphometric Tools for Landmark Data*. Cambridge University Press, Cambridge, 435 p.
- BOOKSTEIN, F. L., P. GUNZ, P. MITTEROECKER, H. PROSSINGER, K. SCHAEFER, AND H. SEIDLER. 2003. Cranial integration in *Homo*: singular warps analysis of the midsagittal plane in ontogeny and evolution. *Journal of Human Evolution*, 44:167-187.
- BOUGHNER, J., S. WAT, V. DIEWERT, N. YOUNG, L. BROWDER, AND B. HALLGRIMSSON. 2008. Short-faced mice and developmental interactions between the brain and the face. *Journal of Anatomy*, 213:646-662.
- BRUNER, E. 2008. Comparing Endocranial Form and Shape Differences in Modern Humans and Neandertals: a Geometric Approach. *PaleoAnthropology*, 2008:93-106.
- BRUNER, E., AND M. RIPANI. 2008. A quantitative and descriptive approach to morphological variation of the endocranial base in modern humans. *American Journal of Physical Anthropology*, 137:30-40.
- BURT, C. 1948. Factor analysis and canonical correlations. *British Journal of Psychology*, 1:95-106.
- CHAPMAN, R. E. 1990. Conventional Procrustes approaches, p. 251-267. *In* F. J. Rohlf and F. L. Bookstein (eds.), *Proceedings of the Michigan Morphometrics Workshop*. University of Michigan Museum of Zoology, Ann Arbor, Michigan.
- CHERNOFF, B., AND P. M. MAGWENE. 1999. Afterword, p. 319-348. *In* E. C. Olson and R. L. Miller, *Morphological Integration*. University of Chicago Press, Chicago.
- CHEVERUD, J. M. 1982. Phenotypic, genetic, and environmental morphological integration in the cranium. *Evolution*, 36:499-516.
- CHEVERUD, J. M. 1988. A comparison of genetic and phenotypic correlations. *Evolution*, 42:958-968.
- CHEVERUD, J. M. 1988. Spatial-analysis in morphology illustrated by rhesus macaque cranial growth and integration. *American Journal of Physical Anthropology*, 75:195-196.

- CHEVERUD, J. M. 1989. A comparative analysis of morphological variation patterns in the Papionines. *Evolution*, 43:1737-1747.
- CHEVERUD, J. M. 1995. Morphological integration in the saddle-back tamarin (*Saguinus fuscicollis*) cranium. *American Naturalist*, 145:63-89.
- CHEVERUD, J. M. 1996a. Developmental integration and the evolution of pleiotropy. *American Zoologist*, 36:44-50.
- CHEVERUD, J. M. 1996b. Quantitative genetic analysis of cranial morphology in the cotton-top (*Saguinus oedipus*) and saddle-back (*S. fuscicollis*) tamarins. *Journal of Evolutionary Biology*, 9:5-42.
- CHEVERUD, J. M. 2004. Modular pleiotropic effects of quantitative trait loci on morphological traits, p. 132-153. *In* G. Schlosser and G. P. Wagner (eds.), *Modularity in development and evolution*. University of Chicago, Chicago.
- CHEVERUD, J. M., J. J. RUTLEDGE, AND W. R. ATCHLEY. 1983. Quantitative genetics of development: Genetic correlations among age-specific trait values and the evolution of ontogeny. *Evolution*, 37:895-905.
- CHEVERUD, J. M., G. P. WAGNER, AND M. M. DOW. 1989. Methods for the comparative analysis of variation patterns. *Systematic Zoology*, 38:201-213.
- CHEVERUD, J. M., E. J. ROUTMAN, AND D. J. IRSCHICK. 1997. Pleiotropic effects of individual gene loci on mandibular morphology. *Evolution*, 51:2006-2016.
- CHEVERUD, J. M., S. E. HARTMAN, J. T. RICHTSMEIER, AND W. R. ATCHLEY. 1991. A quantitative genetic-analysis of localized morphology in mandibles of inbred mice using finite-element scaling analysis. *Journal of Craniofacial Genetics and Developmental Biology*, 11:122-137.
- CHEVERUD, J. M., T. H. EHRICH, T. T. VAUGHN, S. F. KOREISHI, R. B. LINSEY, AND L. S. PLETSCHER. 2004. Pleiotropic effects on mandibular morphology II: Differential epistasis and genetic variation in morphological integration. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution*, 302B:424-435.
- COLE, T.M. III, AND J. RICHTSMEIER. 1998. A simple method for visualization of influential landmarks when using Euclidean distance matrix analysis. *American Journal of Physical Anthropology*, 107:273-283.
- CONNER, J. 2002. Genetic mechanisms of floral trait correlations in a natural population. *Nature*, 420:407-410.
- DE OLIVEIRA, F. B., A. PORTO, AND G. MARROIG. 2009. Covariance structure in the skull of Catarrhini: a case of pattern stasis and magnitude evolution. *Journal of Human Evolution*, 56:417-430.
- DE VISSER, J., J. HERMISSON, G. P. WAGNER, L. MEYERS, H. BAGHERI-CHAICHIAN, J. BLANCHARD, L. CHAO, J. M. CHEVERUD, S. ELENA, AND W. FONTANA. 2003. Perspective: evolution and detection of genetic robustness. *Evolution*, 57:1959-1972.
- DRAGHI, J., AND G. P. WAGNER. 2007. Evolution of evolvability in a developmental model. *Evolution*, 62:301-315.
- DRAKE, A. G., AND C. P. KLINGENBERG. 2010. Large-scale diversification of skull shape in domestic dogs: disparity and modularity. *American Naturalist*, 175:289-301.
- DRYDEN, I. L., AND K. V. MARDIA. 1998. *Statistical Shape Analysis*. John Wiley & Sons, Chichester, U.K. 347 p.
- EBLE, G. 2004. The macroevolution of phenotypic integration, p. 253-273. *In* M. Pigliucci and K. Preston (eds.), *Phenotypic Integration*. Oxford University Press, Oxford.
- EHRICH, T. H., T. T. VAUGHN, S. KOREISHI, R. B. LINSEY, L. S. PLETSCHER, AND J. M. CHEVERUD. 2003. Pleiotropic effects on mandibular morphology I. Developmental morphological integration and differential dominance. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution*, 296B:58-79.
- EMERSON, S. B., AND P. A. HASTINGS. 1998. Morphological correlations in evolution: Consequences for phylogenetic analysis. *The Quarterly Review of Biology*, 73:141-162.
- ESCOUFIER, Y. 1973. Le traitement des variables vectorielles. *Biometrics*, 29:751-760.
- FINK, W. L., AND M. L. ZELDITCH. 1996. Historical patterns of developmental integration in Pirhanas. *American Zoologist*, 36:61-69.
- GONZÁLEZ-JOSÉ, R., I. ESCAPA, W. NEVES, R. CÚNEO, AND H. PUCCIARELLI. 2008. Cladistic analysis of continuous modularized traits provides phylogenetic signals in *Homo* evolution. *Nature*, 453:775-778.
- GOSWAMI, A. 2006a. Cranial modularity shifts during mammalian evolution. *American Naturalist*, 168:270-280.
- GOSWAMI, A. 2006b. Morphological integration in the carnivoran skull. *Evolution*, 60:169-183.
- GOSWAMI, A. 2007a. Cranial integration, phylogeny, and diet in australodelphian marsupials. *PLoSOne*, 2:e995.
- GOSWAMI, A. 2007b. Cranial modularity and sequence heterochrony in mammals. *Evolution & Development*, 9:290-298.
- GOSWAMI, A., AND P. D. POLLY. 2010a. The influence of character correlations of phylogenetic analyses: a case study of the carnivoran cranium, p. 141-164. *In* A. Goswami and A. Friscia (eds.), *Carnivoran Evolution: New Views on Phylogeny, Form, and Function*. Cambridge University Press, Cambridge.
- GOSWAMI, A., AND P. D. POLLY. 2010b. The influence of modularity on cranial morphological diversity in Carnivora and Primates (Mammalia; Placentalia). *PLoS One*, 5:e9517.
- GOSWAMI, A., V. WEISBECKER, AND M. R. SÁNCHEZ-VILLAGRA. 2009. Developmental modularity and the marsupial-placental dichotomy. *Journal of Experimental Zoology*, 312B:186-195.

- HALLGRÍMSSON, B., K. WILLMORE, AND B. HALL. 2002. Canalization, developmental stability, and morphological integration in primate limbs. *American Journal of Physical Anthropology*, 119:131-158.
- HALLGRÍMSSON, B., C. DORVAL, M. ZELDITCH, AND R. GERMAN. 2004a. Craniofacial variability and morphological integration in mice susceptible to cleft lip and palate. *Journal of Anatomy*, 205:501.
- HALLGRÍMSSON, B., K. WILLMORE, C. DORVAL, AND D. M. L. COOPER. 2004b. Craniofacial variability and modularity in macaques and mice. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution*, 302B:207-225.
- HALLGRÍMSSON, B., H. JAMNICZKY, N. M. YOUNG, C. ROLIAN, T. E. PARSONS, J. C. BOUGHNER, AND R. S. MARCUCIO. in press. Deciphering the palimpsest: studying the relationship between morphological integration and phenotypic covariation. *Evolutionary Biology*.
- HANSEN, T., AND D. HOULE. 2008. Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology*, 21:1201-1219.
- HERMISSE, J., AND G. P. WAGNER. 2004. Evolution of phenotypic robustness, p. 47-70. *In* E. Jen (ed.), *Robust Design: A Repertoire from Biology, Ecology, and Engineering Case Studies*. Oxford University Press, Oxford.
- HOTELLING, H. 1936. Relations between two sets of variates. *Biometrika*, 28:321.
- JOLICOEUR, P., AND J. E. MOSIMANN. 1960. Size and shape variation in the painted turtle. A principal component analysis. *Growth*, 24:339-354.
- JOLLIFE, I. T. 2002. *Principal Components Analysis*. Springer-Verlag, New York, 487 p.
- JONES, A., S. ARNOLD, AND R. BÜRGER. 2007. The mutation matrix and the evolution of evolvability. *Evolution*, 61:727-745.
- KLINGENBERG, C. P. 2003. Developmental instability as a research tool: using patterns of fluctuating asymmetry to infer the developmental origins of morphological integration. *Developmental instability: causes and consequences*:427-442.
- KLINGENBERG, C. P. 2004. Integration, modules, and development: molecules to morphology to evolution, p. 213-230. *In* M. Pigliucci and K. Preston (eds.), *Phenotypic Integration*. Oxford University Press, Oxford.
- KLINGENBERG, C. P. 2005. Developmental constraints, modules, and evolvability, p. 219-247. *In* B. Hallgrímsson and B. K. Hall (eds.), *Variation*. Academic Press, San Diego.
- KLINGENBERG, C. P. 2008a. Morphological integration and developmental modularity. *Annual Review of Ecology, Evolution and Systematics*, 39: 115-132.
- KLINGENBERG, C. P. 2008b. MorphoJ. http://www.flywings.org.uk/MorphoJ_page.htm. Faculty of Life Sciences, University of Manchester, UK.
- KLINGENBERG, C. P. 2009. Morphometric integration and modularity in configurations of landmarks: tools for evaluating a priori hypotheses. *Evolution & Development*, 11:405-421.
- KLINGENBERG, C. P., AND S. D. ZAKLAN. 2000. Morphological integration between developmental compartments in the *Drosophila* wing. *Evolution*, 53:358-375.
- KLINGENBERG, C. P., AND L. J. LEAMY. 2001. Quantitative genetics of geometric shape in the mouse mandible. *Evolution*, 55:2342-2352.
- KLINGENBERG, C. P., A. V. BADYAEV, S. M. SOWRY, AND N. J. BECKWITH. 2001a. Inferring developmental modularity from morphological integration: analysis of individual variation and asymmetry in bumblebee wings. *American Naturalist*, 157:11-23.
- KLINGENBERG, C. P., L. J. LEAMY, E. J. ROUTMAN, AND J. M. CHEVERUD. 2001b. Genetic architecture of mandible shape in mice: Effects of quantitative trait loci analyzed by geometric morphometrics. *Genetics*, 157:785-802.
- KLINGENBERG, C., M. BARLUENGA, AND A. MEYER. 2002. Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. *Evolution*, 56:1909-1920.
- KLINGENBERG, C. P., K. MEBUS, AND J. C. AUFRAY. 2003. Developmental integration in a complex morphological structure: how distinct are the modules in the mouse mandible? *Evolution & Development*, 5:522-531.
- KLINGENBERG, C. P., L. J. LEAMY, AND J. M. CHEVERUD. 2004. Integration and modularity of quantitative trait locus effects on geometric shape in the mouse mandible. *Genetics*, 166:1909-1921.
- LAURITZEN, S. L. 1996. *Graphical Models*. Clarendon Press, Oxford, x, 298 p. p.
- LEAMY, L. J., E. J. ROUTMAN, AND J. M. CHEVERUD. 1999. Quantitative trait loci for early- and late-developing skull characters in mice: A test of the genetic independence model of morphological integration. *American Naturalist*, 153:201-214.
- LEAMY, L. J., E. J. ROUTMAN, AND J. M. CHEVERUD. 2002. An epistatic genetic basis for fluctuating asymmetry of mandible size in mice. *Evolution*, 56:642-653.
- LEGENBRE, P., AND L. LEGENBRE. 1998. *Numerical Ecology*, Elsevier Science, Amsterdam, 870 p.
- LELE, S., AND J. RICHTSMEIER. 1991. Euclidean distance matrix analysis: a coordinate-free approach for comparing biological shapes using landmark data. *American Journal of Physical Anthropology*, 86:415-427.

- LELE, S., AND J. RICHTSMIEIER. 2001. An Invariant Approach to Statistical Analysis of Shapes. CRC Press, Boca Raton, FL, 308 p.
- LELE, S. R. 1999. Invariance and morphometrics: a critical appraisal of statistical techniques for landmark data, p. 325-336. *In* M. Chaplain, G. Singh, and J. McLachlan (eds.), *On Growth and Form: Spatiotemporal Pattern Formation in Biology*. John Wiley and Sons, Chichester, UK.
- LIEBERMAN, D. E., C. F. ROSS, AND M. J. RAVOSA. 2000a. The primate cranial base: Ontogeny, function, and integration. *Yearbook of Physical Anthropology*, 43:117-169.
- LIEBERMAN, D. E., O. M. PEARSON, AND K. M. MOWBRAY. 2000b. Basicranial influence on overall cranial shape. *Journal of Human Evolution*, 38:291-315.
- MAGWENE, P. M. 2001. New tools for studying integration and modularity. *Evolution*, 55:1734-1745.
- MAGWENE, P. M. 2008. Using correlation proximity graphs to study phenotypic integration. *Evolutionary Biology*, 35:191-198.
- MAGWENE, P. M. 2009. Statistical methods for studying modularity: A reply to Mitteroecker and Bookstein. *Systematic Biology*, 58:146-149.
- MANTEL, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27:209-220.
- MÁRQUEZ, E. J. 2008. A statistical framework for testing modularity in multidimensional data. *Evolution*, 62:2688-2708.
- 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.
- MARROIG, G., AND J. M. CHEVERUD. 2004. Cranial evolution in sakis (*Pithecia, platyrrhini*) I: Interspecific differentiation and allometric patterns. *American Journal of Physical Anthropology*, 125:266-278.
- MARROIG, G., AND J. CHEVERUD. 2005. Size as a line of least evolutionary resistance: diet and adaptive morphological radiation in new world monkeys. *Evolution*, 59:1128-1142.
- MARROIG, G., M. VIVO, AND J. M. CHEVERUD. 2004. Cranial evolution in sakis (*Pithecia, Platyrrhini*) II: Evolutionary processes and morphological integration. *Journal of Evolutionary Biology*, 17:144-155.
- MCGUIGAN, K., S. CHENOWETH, AND M. BLOWS. 2005. Phenotypic divergence along lines of genetic variance. *American Naturalist*, 165:32-43.
- MITTEROECKER, P., AND F. BOOKSTEIN. 2007. The conceptual and statistical relationship between modularity and morphological integration. *Systematic Biology*, 56:818-836.
- MITTEROECKER, P., AND F. BOOKSTEIN. 2009. Examining modularity via partial correlations: a rejoinder to a comment by Paul Magwene. *Systematic Biology*, 58:346-348.
- MITTEROECKER, P., P. GUNZ, M. BERNHARD, K. SCHAEFER, AND F. L. BOOKSTEIN. 2004. Comparison of cranial ontogenetic trajectories among great apes and humans. *Journal of Human Evolution*, 46:679-697.
- MONTEIRO, L., V. BONATO, AND S. DOS REIS. 2005. Evolutionary integration and morphological diversification in complex morphological structures: mandible shape divergence in spiny rats (Rodentia, Echimyidae). *Evolution & Development*, 7:429-439.
- MOSS, M. O., AND R. W. YOUNG. 1960. A functional approach to craniology. *American Journal of Physical Anthropology*, 18:281-291.
- NIJHOUT, H. 2002. The nature of robustness in development. *BioEssays*, 24:553-563.
- O'KEEFE, F. R., AND P. J. WAGNER. 2001. Inferring and testing hypotheses of cladistic character dependence by using character compatibility. *Systematic Biology*, 50:657-675.
- OLSON, E. C., AND R. L. MILLER. 1958. *Morphological Integration*. University of Chicago Press, Chicago, 376 p.
- PALMER, A., AND C. STROBECK. 1986. Fluctuating asymmetry: Measurement, analysis, patterns. *Annual Review of Ecology and Systematics*, 17:391-421.
- PAVLICEV, M., J. M. CHEVERUD, AND G. P. WAGNER. 2009. Measuring morphological integration using eigenvalue variance. *Evolutionary Biology*, 36:157-170.
- PHILLIPS, P. C., AND S. J. ARNOLD. 1999. Hierarchical comparison of genetic variance-covariance matrices. I. Using the Flury hierarchy. *Evolution*, 53:1506-1515.
- PIELOU, E. C. 1984. Probing multivariate data with random skewers: a preliminary to direct gradient analysis. *Oikos*, 42:161-165.
- PIGLIUCCI, M., AND K. PRESTON. 2004. *Phenotypic Integration*. Oxford University Press, Oxford, 464 p.
- POLLY, P. D. 2001. On morphological clocks and paleophylogeography: towards a timescale for Sorex hybrid zones. *Genetica*, 112:339-357.
- POLLY, P. D. 2004. On the simulation of morphological shape: multivariate shape under selection and drift, *Palaeontologia Electronica*, 7:7A.
- POLLY, P. D. 2005. Development and phenotypic correlations: the evolution of tooth shape in *Sorex araneus*. *Evolution & Development*, 7:29-41.
- POLLY, P. D. 2008. Developmental dynamics and G-matrices: Can morphometric spaces be used to model phenotypic evolution? *Evolutionary Biology*, 35:83-96.

- RAUP, D. 1975. Taxonomic diversity estimation using rarefaction. *Paleobiology*, 333-342.
- RENAUD, S., J. AUFRAY, AND J. MICHAUX. 2006. Conserved phenotypic variation patterns, evolution along lines of least resistance, and departure due to selection in fossil rodents. *Evolution*:1701-1717.
- RICE, S. 1998. The evolution of canalization and the breaking of Von Baer's laws: Modeling the evolution of development with epistasis. *Evolution*, 52:647-656.
- RICE, S. H. 2008. The G-matrix as one piece of the phenotypic evolution puzzle. *Evolutionary Biology*, 35:106-107.
- RICHTSMEIER, J., AND S. LELE. 1993. A coordinate-free approach to the analysis of growth patterns: models and theoretical considerations. *Biological Reviews*, 68:381-381.
- RICHTSMEIER, J. T., V. B. DELEON, AND S. R. LELE. 2002. The promise of geometric morphometrics. *American Journal of Physical Anthropology*, 119 (S35):63-91.
- ROFF, D. 1995. The estimation of genetic correlations from phenotypic correlations: a test of Cheverud's conjecture. *Heredity*, 74:481-490.
- ROFF, D. 2000. The evolution of the G matrix: selection or drift? *Heredity*, 84:135-142.
- ROFF, D., AND T. MOUSSEAU. 2005. The evolution of the phenotypic covariance matrix: evidence for selection and drift in *Melanoplus*. *Journal of Evolutionary Biology*, 18:1104-1114.
- ROHLF, F. J. 1990. Rotational fit Procrustes methods. University of Michigan Museum of Zoology Special Publications, 2:227-236.
- ROHLF, F. J. 1993. Relative warp analysis and an example of its application to mosquito wings., p. 131-159. In L. F. Marcus, E. Bello, and A. Garcia-Valdecasas (eds.), *Contributions to Morphometrics*. Museo Nacional de Ciencias Naturales, Madrid.
- ROHLF, F. J. 2000. On the use of shape spaces to compare morphometric methods. *Hystrix*, 11:9-25.
- ROHLF, F. J. 2003. Bias and error in estimates of mean shape in geometric morphometrics. *Journal of Human Evolution*, 44:665-683.
- ROHLF, F. J., AND D. SLICE. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, 39:40-59.
- ROHLF, F. J., AND M. CORTI. 2000. Use of two-block partial least-squares to study covariation in shape. *Systematic Biology*, 49:740-753.
- SADLEIR, R. W., AND P. J. MAKOVICKY. 2008. Cranial shape and correlated characters in crocodile evolution. *Journal of Evolutionary Biology*, 21:1578-1596.
- SANDERS, H. 1968. Marine benthic diversity: a comparative study. *American Naturalist*, 102:243.
- SCHLOSSER, G., AND G. P. WAGNER. 2004. *Modularity in development and evolution*. University of Chicago Press, Chicago, 600 p.
- SCHLUTER, D. 1996. Adaptive radiation along genetic lines of least resistance. *Evolution*, 50:1766-1774.
- SCHWENK, K. 2001. Functional units and their evolution, p. 165-198. In G. P. Wagner (ed.), *The character concept in evolutionary biology*. Academic Press, San Diego, CA.
- SNEATH, P., AND R. SOKAL. 1973. *Numerical Taxonomy: the Principles and Practice of Numerical Classification*. W.H. Freeman, San Francisco, 588 p.
- STEBBING, S. 1997. Phylogenetic analysis of phenotypic covariance structure. I. Contrasting results from matrix correlation and common principal component analysis. *Evolution*, 51:571-586.
- STEBBING, S., P. PHILLIPS, AND D. HOULE. 2002. Comparative quantitative genetics: evolution of the G matrix. *Trends in Ecology & Evolution*, 17:320-327.
- STRAIT, D. S. 2001. Integration, phylogeny, and the hominid cranial base. *American Journal of Physical Anthropology* 114:273-297.
- STREISSGUTH, A., F. BOOKSTEIN, P. SAMPSON, AND H. BARR. 1993. The enduring effects of prenatal alcohol exposure on child development: Birth through seven years, a partial least squares solution. University of Michigan Press, Ann Arbor.
- VAN DER KLAUW, C. J. 1948-1952. Size and position of the functional components of the skull. *Archives Neerlandaises de Zoologie*, 9:1-559.
- VAN VALEN, L. 1962. A study of fluctuating asymmetry. *Evolution*, 16:125-142.
- VERMEIJ, G. J. 1973. Adaptation, versatility, and evolution. *Systematic Zoology*, 22:466-477.
- WADDINGTON, C. 1942. Canalization of development and the inheritance of acquired characters. *Nature*, 150:563-565.
- WAGNER, A. 2008. Robustness and evolvability: a paradox resolved. *Proceedings of the Royal Society of London, Series B*, 275:91.
- WAGNER, G. P. 1988. The influence of variation and of developmental constraints on the rate of multivariate phenotypic evolution. *Journal of Evolutionary Biology*, 1:45-66.
- WAGNER, G. P. 1995. Adaptation and the modular design of organisms. *Advances in Artificial Life*, 929:317-328.
- WAGNER, G. P. 1996. Homologues, natural kinds and the evolution of modularity. *American Zoologist*, 36:36-43.
- WAGNER, G. P., AND L. ALTENBERG. 1996. Perspective: Complex adaptations and the evolution of evolvability. *Evolution*, 50:967-976.

- WAGNER, G. P., AND J. G. MEZEY. 2004. The role of genetic architecture constraints in the origin of variational modularity, p. 338-358. *In* G. Schlosser and G. P. Wagner (eds.), *Modularity in Development and Evolution*. University of Chicago Press, Chicago.
- WAGNER, G. P., G. BOOTH, AND H. BAGHERI-CHAICHIAN. 1997. A population genetic theory of canalization. *Evolution*, 51:329-347.
- WAGNER, G. P., M. PAVLICEV, AND J. M. CHEVERUD. 2007. The road to modularity. *Nature Genetics*, 8:921-931.
- WAGNER, P. J. 1998. A likelihood approach for evaluating estimates of phylogenetic relationships among fossil taxa. *Palaeobiology*, 24:430-449.
- WARD JR, J. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58:236-244.
- WEBSTER, M., AND M. L. ZELDITCH. 2008. Integration and regulation of developmental systems in trilobites. *Cuadernos del Museo Geominero*, 9:427-433.
- WHITTAKER, J. 1990. *Graphical Models in Applied Multivariate Statistics*. Wiley, Chichester U.K., 448 p.
- YOUNG, N., AND B. HALLGRÍMSSON. 2005. Serial homology and the evolution of mammalian limb covariation structure. *Evolution*, 59:2691-2704.
- 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. CARMICHAEL. 1989a. Growth and intensity of integration through postnatal growth in the skull of *Sigmodon fulviventer*. *Journal of Mammalogy*, 70:477-484.
- ZELDITCH, M. L., AND A. C. CARMICHAEL. 1989b. Ontogenetic variation in patterns of developmental and functional integration in skulls of *Sigmodon fulviventer*. *Evolution* 43:814-824
- ZELDITCH, M. L., AND R. A. MOSCARELLA. 2004. Form, function and life history: spatial and temporal dynamics of integration, p. 274-301 *In* M. Pigliucci and K. Preston (eds.), *Phenotypic Integration*. Oxford University Press, Oxford
- ZELDITCH, M. L., H. D. SHEETS, AND W. L. FINK. 2001. The spatial complexity and evolutionary dynamics of growth, p. 145-194 *In* M. L. Zelditch (ed.), *Beyond Heterochrony: The Evolution of Development*. Wiley-Liss, Inc., New York.
- ZELDITCH, M. L., D. L. SWIDERSKI, H. D. SHEETS, AND W. L. FINK. 2004. *Geometric Morphometrics for Biologists: A Primer*. Elsevier Academic Press, San Diego, 416 p.
- ZELDITCH, M. L., J. MEZEY, H. SHEETS, B. LUNDRIGAN, AND T. GARLAND. 2006. Developmental regulation of skull morphology II: Ontogenetic dynamics of covariance. *Evolution and Development*, 8:46-60.
- ZELDITCH, M. L., A. R. WOOD, R. M. BONETT, AND D. L. SWIDERSKI. 2008. Modularity of the rodent mandible: Integrating bones, muscles, and teeth. *Evolution & Development*, 10:756-768.
- ZELDITCH, M. L., A. WOOD, AND D. SWIDERSKI. 2009. Building developmental integration into functional systems: function-induced integration of mandibular shape. *Evolutionary Biology*, 36:71-87.

