

# A STATISTICAL FRAMEWORK FOR TESTING MODULARITY IN MULTIDIMENSIONAL DATA

Eladio J. Márquez<sup>1,2</sup>

<sup>1</sup>*Department of Ecology & Evolutionary Biology, University of Michigan, Ann Arbor Michigan 48109*

<sup>2</sup>*E-mail: emarquez@umich.edu*

Received April 24, 2008

Accepted July 15, 2008

Modular variation of multivariate traits results from modular distribution of effects of genetic and epigenetic interactions among those traits. However, statistical methods rarely detect truly modular patterns, possibly because the processes that generate intramodular associations may overlap spatially. Methodologically, this overlap may cause multiple patterns of modularity to be equally consistent with observed covariances. To deal with this indeterminacy, the present study outlines a framework for testing a priori hypotheses of modularity in which putative modules are mathematically represented as multidimensional subspaces embedded in the data. Model expectations are computed by subdividing the data into arrays of variables, and intermodular interactions are represented by overlapping arrays. Covariance structures are thus modeled as the outcome of complex and nonorthogonal intermodular interactions. This approach is demonstrated by analyzing mandibular modularity in nine rodent species. A total of 620 models are fit to each species, and the most strongly supported are heuristically modified to improve their fit. Five modules common to all species are identified, which approximately map to the developmental modules of the mandible. Within species, these modules are embedded within larger “super-modules,” suggesting that these conserved modules act as building blocks from which covariation patterns are built.

**KEY WORDS:** Developmental modules, dimensionality, geometric morphometrics, morphological integration, rodent mandible, variational modules.

It is often hypothesized that the genotype–phenotype map has a modular structure, meaning that pleiotropic effects tend to be restricted to discrete subsets of phenotypic traits (Wagner and Altenberg 1996). These highly integrated subsets, termed “variational modules” (Wagner and Mezey 2004), are empirically recognized by strong statistical associations among their component traits, in contrast to their weak associations with traits in other modules (Olson and Miller 1958; Cheverud 1982; Magwene 2001; Wagner and Mezey 2004; Klingenberg 2005). Modularity of the genotype–phenotype map has become an important focus for empirical and theoretical research, largely because it is viewed as a condition for evolvability, i.e., the ability to produce selectively useful variation (Wagner and Altenberg 1996; Hansen 2006; Wagner et al. 2007; Draghi and Wagner 2008). The prevailing theory is that pleiotropy evolves under natural selection to match the patterns of functional and developmental interdepen-

dencies among traits (Riedl 1977; Cheverud 1982; Wagner 1988; Ehrlich et al. 2003; Pavlicev et al. 2008). This thesis has stimulated numerous empirical studies that seek to determine whether patterns of statistical association do match predictions derived from developmental and functional theories (Olson and Miller 1958; Berg 1960; Cheverud 1982; Zelditch and Carmichael 1989; Marroig and Cheverud 2001; Armbruster et al. 2004; Monteiro et al. 2005). As most studies demonstrate, however, the empirical detection of modules is rarely a straightforward affair, primarily because covariance matrices rarely, if ever, show a truly modular structure.

Typically, modularity is expected to contribute only to statistical associations within modules; but covariances between modules are rarely equal to zero (Klingenberg et al. 2003; Armbruster et al. 2004; Monteiro et al. 2005). This creates a situation in which more than one pattern of modularity may be supported by the data

(i.e., model indeterminacy), which must be addressed by explicitly modeling intra- and intermodular covariation. In this sense, intermodular associations are sometimes explained in terms of general factors that span all modules (i.e., overall integration; Wright 1932; Magwene 2001; Mitteroecker and Bookstein 2007, 2008), and are factored out to highlight intramodular associations. Other approaches do not attempt to explain these associations at all, treating them instead as random deviations from the expectation that intermodular associations equal zero or an arbitrarily low value (Cheverud 1982; Ackermann and Cheverud 2000; Marroig and Cheverud 2001; Klingenberg et al. 2003; Monteiro et al. 2005). Yet, as suggested by patterns of between-module QTL effects (Cheverud et al. 1997; Mezey et al. 2000; Ehrich et al. 2003; Cheverud 2004; Albertson et al. 2005), local integration between modules may be both biologically real and important in the development and evolution of modularity (see Wagner and Mezey 2004). Most importantly, intermodular interactions can be expected if it is considered that modules result from the cumulative effects of complex, spatially overlapping processes. Such complexity should favor the production of a “collage” of overlapping, local modules that may appear to be globally integrated (Cowley and Atchley 1992; Hallgrímsson et al. 2007; Boughner and Hallgrímsson 2008).

This view of covariation patterns as the outcome of complex interactions between localized but overlapping factors has important implications for the conceptual and empirical delimitation of modules. Theoretically, an observed covariance structure is expected to reflect the cumulative effect of a number of embedded dimensions (Orr 2000; Mezey and Houle 2005; Hine and Blows 2006). Each of these dimensions corresponds in turn to a specific set of genetic, developmental, or functional interactions (Wagner 1988), and these can vary and therefore evolve semi-independently from other, similar processes (Raff 1996; Beldade et al. 2002; Hallgrímsson et al. 2007). Therefore, covariance structures can be hypothesized to diverge even if they maintain a conserved intrinsic structure in that associations among modules may readily evolve even if the modules themselves are conserved (Klingenberg 2005).

Finding those conserved modules within a dynamic structure of intermodular associations presents a real methodological challenge (Zelditch et al. 1990; Hallgrímsson et al. 2007). Methods such as Common Principal Components Analysis might seem well-suited for this problem because of their emphasis on finding dimensions shared by multiple patterns of covariation (Klingenberg and Zimmermann 1992; Stepan 1997; Phillips and Arnold 1999); but such methods do not define those shared dimensions in terms of homologous developmental processes or modules, in that ascribing such meanings to shared dimensions usually requires a priori hypotheses of modularity (Zelditch et al. 1990). Combining tests of mechanistic hypotheses and interspe-

cific comparisons of the models supported by multiple species may thus be necessary if we hope to understand the developmental processes that produce complex covariance structures and the impact of these processes on evolutionary divergence (Riska 1986; Hall 2003; Armbruster et al. 2004).

The objective of the present study is to present a statistical framework for testing hypotheses of variational modularity. The approach outlined herein assesses the fit of competing models derived from developmental and functional theories. In contrast to other methods based on the same general premise (Cheverud 1982; Zelditch et al. 1990; Monteiro et al. 2005), the one presented herein treats each module as a multidimensional array throughout the entire analysis, on the grounds that most complex phenotypic traits, and hence modules, are inherently multidimensional objects (Márquez and Knowles 2007). This treatment of modules as high-dimensional structures allows for separate measurement of intra- and intermodular integration (see Mezey et al. 2000). In addition, the approach allows for testing hypotheses that predict complex overlapping effects between modules, in that models are built such that the variation of intramodular dimensions contributes to the variation of each overlapping module. Treating module overlap in this way allows for explicitly modeling the causes of intermodular integration in terms of the spatial overlap of the responsible processes. Finally, by maintaining the multidimensional identity of modules, it is possible to use comparative analyses to search for the modules shared by several species. The joint analysis of conserved and divergent modules (or components thereof) provides a link between the mechanistic explanations for the origin of covariation, and the patterns of divergence of covariance matrices.

The methodology is demonstrated by a case study of a clade of Neotropical rodents, chosen for their remarkable ecological diversity (Voss 1988). This ecological diversity predicts functional diversity of the mandible (Smith 1993; Tokita et al. 2007), the model system of choice in many studies of modularity and integration (Leamy et al. 1999; Klingenberg et al. 2003; Young and Badyaev 2006). The mandible has become a favored model system largely because it is a single bone with a complex structure that develops from the integration of six mesenchymal condensations (Atchley and Hall 1991; Hall and Miyake 2000), the developmental modules of the mandible (Hall 2003). In this study, competing theories that predict integration of these developmental modules into larger variational modules are simultaneously tested. The complexity of these models ranges from the case of nonoverlapping modules to cases that predict the cumulative effect of multiple overlapping processes. The results of the analysis suggest that no single process can account for the patterns of modularity observed within any species and that overlapping of nonhierarchical modules appears to be uncommon. Comparative analyses discern substantial diversity of patterns of modularity, suggesting that shifts in the relative importance of individual developmental

processes have occurred in the radiation of these rodents. Nonetheless, the comparative analysis also reveals phylogenetically conserved modules that closely map to the developmental modules of the mandible (Hall and Miyake 2000; Hall 2003).

## Materials and Methods

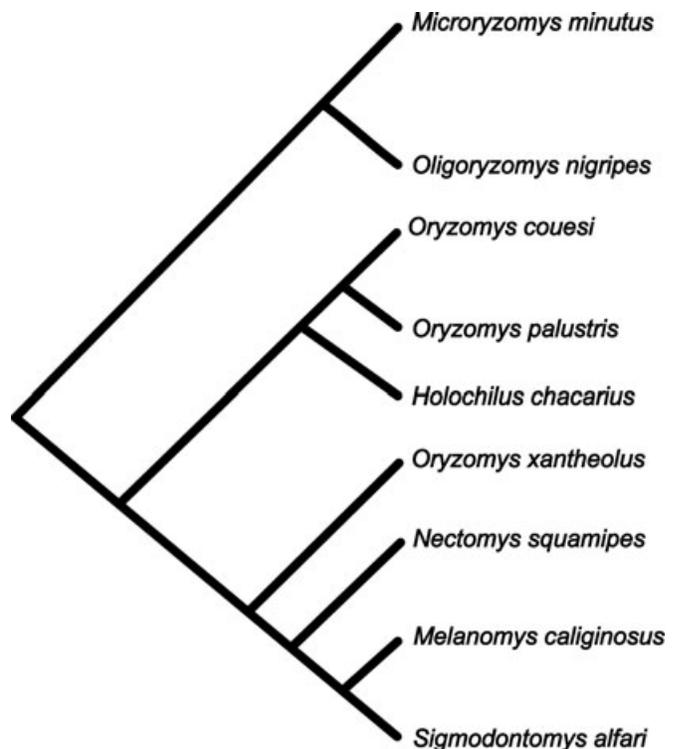
### SAMPLING

The species sampled for present analyses belong to the oryzomyine clade of sigmodontine rodents (Table 1), which have, in the relatively short period of time that has elapsed since the Great American Faunal Exchange (Marshall et al. 1982), experienced a remarkable ecological radiation and expansion throughout South, Central, and Southern North American (Carleton 1973; Voss 1988). Nine species were chosen based on phylogenetic proximity (Weksler 2003, 2006) and to ensure an appropriate representation of the ecological diversity of the clade (Table 1). Figure 1 shows the topology of the portion of the oryzomyine phylogeny encompassing the species sampled in this study (Weksler 2006). Although relevant ecological features such as dietary preferences are not known in detail for most species, their wide distribution, ranging from habitats as distinct as the Peruvian desert (e.g., *Oryzomys xantheolus*; Guabloche et al. 2002) to rainforest streams (e.g., *Nectomys squamipes*; Hershkovitz 1944), strongly suggests that these species span a wide dietary spectrum (see Supporting Table S1).

Samples were drawn from museum specimens (see Supporting Table S1). Because this study was focused on estimation of population (co)variation patterns that could be explained in terms of developmental and functional interactions, it was important to control for extraneous sources of variation (e.g., geographical location, collection date, sex), which might also induce phenotypic covariances, therefore obscuring patterns of modularity. Specimens were thus chosen to maximize sample size while limiting geographical and temporal variation as much as possible. To determine whether to pool samples from multiple geographical lo-

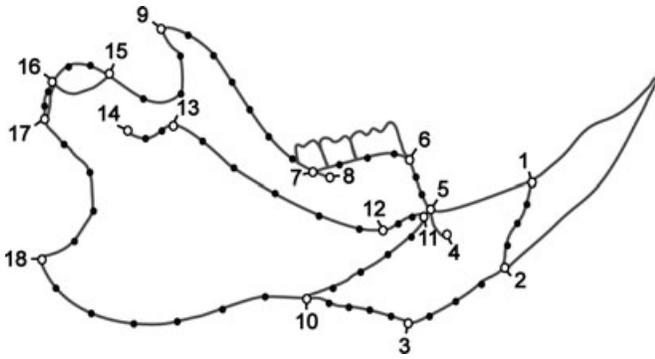
**Table 1.** List of oryzomyine species and sample sizes ( $N$ ) used in this study.

Sampled species	$N$
<i>Holochilus chacarius</i>	64
<i>Melanomys caliginosus</i>	52
<i>Microryzomys minutus</i>	69
<i>Nectomys squamipes</i>	77
<i>Oligoryzomys nigripes</i>	61
<i>Oryzomys couesi</i>	64
<i>Oryzomys palustris</i>	70
<i>Oryzomys xantheolus</i>	50
<i>Sigmodontomys alfari</i>	39



**Figure 1.** Phylogenetic relationships among the oryzomyine species included in this study. Modified from Weksler (2006).

cations, collection dates, or sexes prior to computing phenotypic covariances, the following procedure was carried out to test for homogeneity of covariance matrices among these samples: first, similarity between covariance matrices (e.g., males vs. females of a species) was measured as the angle between the subspaces spanned by each sample (Zelditch et al. 2006). In practice, each of these subspaces is defined by the first  $u$  eigenvectors of a sample (i.e., Principal Components), and in this study  $u$  was chosen so that each subspace spanned at least 95% of the variation in a sample. Whenever  $u$  differed between the two samples being compared (e.g., due to differences in sample size), the smallest value was used for both. Expectations from the null hypothesis that covariance matrices were no more different than expected by chance were then obtained by permuting species datasets by randomly assigning specimens to each sex, geographical location, and capture date while maintaining the original sample sizes. Angles between subspaces were computed for each of 1000 random permutations produced in this way, and  $P$ -values for the null hypothesis were computed by dividing the number of permutations in which the angle equaled or exceeded the original value by the total number of permutations. Most of these comparisons supported pooling of samples ( $P > 0.05$ ; see Supporting Table S2). Only the samples comprising *Oryzomys couesi* captured in Guatemala ( $N = 64$ ) and Mexico ( $N = 18$ ) differed significantly and so the latter sample was removed from present analyses. Final sample sizes



**Figure 2.** Diagrammatic representation of the rodent mandible indicating the landmarks (open circles, labeled) and semilandmarks (closed circles, unlabeled) sampled for this study.

range from 39 for *Sigmodontomys alfari* to 77 for *N. squamipes* (Table 1). Only adult specimens, as determined by tooth eruption and wear, were used in the present study.

Digital images were acquired from right-side mandibles in lateral (labial) view except when that side was severely damaged. Specimens were placed in standard orientation and landmarks and points along curves (semilandmarks) were digitized on a tablet PC using TpsDig2 (Rohlf 2006). A total of 18 landmarks and 51 semilandmarks were sampled (Fig. 2). Landmarks were superimposed by the conventional Procrustes superimposition (Rohlf and Slice 1990); semilandmarks were superimposed by allowing them to slide along curves bounded by landmarks to minimize the Procrustes distance among individuals (Bookstein 1997). Superimposition of semilandmarks was done in Semiland6 (Sheets 2002a); allometric variation was removed by calculating residuals from a regression of shape on centroid size using Standard 6 (Sheets 2002b).

The use of semilandmarks and superimposition reduce the dimensionality of a sample so that the data occupy fewer dimensions than there are ( $x, y$ ) coordinates. Semilandmarks supply only one dimension because they can vary only in the direction perpendicular to the curve (Bookstein 1997) and four dimensions are lost from the entire configuration due to scaling, translation, and rotation (Bookstein 1991). This discrepancy between the dimensionality of the data and the number of coordinates is not problematic for this analysis because the comparisons between hypothetical and observed covariance matrices are based on an integral metric (see below) that uses all the information in covariance matrices to produce a single scalar value. The results should therefore be invariant to geometric transformations of these matrices, differing only by scale. Therefore, ordinary Procrustes residuals (Dryden and Mardia 1998) were used to compute covariance matrices in all present analyses. The method produces the same results whether covariances are derived from Procrustes

residuals or Partial Warp scores so long as they are properly scaled.

### TESTING PATTERNS OF VARIATIONAL MODULARITY: GENERAL APPROACH

In the present study, patterns of variational modularity are examined by testing alternative a priori models, each of which hypothesizes a distinct modular structure caused by specific functional or developmental mechanisms. In practice, the models are represented by a series of partitions, each of which delimits a region sampled by subsets of landmarks and semilandmarks. Each anatomical subset of coordinates corresponds to a putative module and so ought to be characterized by high internal morphological integration and low integration with other such modules. The models selected for testing represent a diversity of processes and events spanning morphogenesis through postweaning growth and remodeling, including processes such as mesenchymal condensation (Atchley and Hall 1991; Hall and Miyake 2000) and effects of masticatory muscles on bone deposition and remodeling (Herring 1993). These hypotheses are a far from exhaustive list but nonetheless cover a substantial proportion of the developmental and functional processes capable of affecting covariation patterns in the mouse mandible.

### MODELS

A total of 13 models are examined in this study ( $H_0$ – $H_{12}$ ), including a “null” model ( $H_0$ ), which states that there is no integration or modularity (i.e., all covariances equal zero). Each model specifies modules as nonoverlapping sets of landmarks (see Table 2 and Fig. 3) and they are analogous to hypotheses of integration and modularity often tested in similar studies (Klingenberg et al. 2003) in that they specify a simple structure (Mitteroecker and Bookstein 2007). This simple structure can be termed “strict modularity,” inasmuch as all modules are statistically independent of all others. Even though these individual modules are orthogonal by design, they are not statistically or biologically equivalent to the orthogonal vectors produced by eigenanalysis techniques (e.g., PCA, CPCA) because those eigenvectors are computed to maximize explained variance, and thus will only match a strictly modular structure if the (orthogonal) modules per se account for decreasing proportions of the variance of the data.

The modules expected from developmental and functional theories can be conveniently represented as nonoverlapping, orthogonal units, but these hypotheses are not mutually exclusive and it is likely that multiple processes will structure covariances, producing more complex structures that can be appropriately modeled by combining the modules associated to multiple hypotheses. For that reason, all possible nonhierarchical combinations of the modules proposed by the original 12 hypotheses of

**Table 2.** A priori developmental and functional modules of modularity tested in this study. Modules correspond to regions of the mandible hypothesized to be affected by a common process. Partitions are: {1} anterior (rostral) portion of incisor alveolus; {2} molar alveolus; {3} coronoid process; {4} posterior portion of incisor alveolus and attachment site of lateral masseter muscles, decomposable into partitions {4<sup>+</sup>} and {4<sup>-</sup>}; {5} condyloid process; {6} angular process. See Figure 3 for a pictorial representation of each partition.

Model	Description
H <sub>0</sub> : No modules	“Null” model, predicting absence of modular structure; all covariances are hypothesized to be zero.
H <sub>1</sub> : {1}{2}{3}{4}{5}{6} H <sub>2</sub> : {1,4}{2}{3}{5}{6}	Partitions map to mesenchyme condensations, the developmental modules of the mandible. Two versions are distinguished, based on whether partition {4} is associated to the condensation that gives rise to ramal bone (H <sub>1</sub> ) or to the incisor alveolus (H <sub>2</sub> ) (Atchley and Hall 1991; Hall and Miyake 2000; Hall 2003).
H <sub>3</sub> : {1,2,4}{3,5,6}	Partitions distinguish the processes of the mandible from the ramus and alveoli, which undergo endochondral and membranous ossification, respectively. This model also addresses timing of ossification as a source of integration, because membranous ossification precedes endochondral ossification (Kaufman and Bard 1999)
H <sub>4</sub> : {1,4 <sup>+</sup> }{2}{5}{3,4 <sup>-</sup> ,6}	Molar and incisor alveoli defined as separate partitions to represent developmental and functional interactions with their respective teeth (Boughner and Hallgrímsson 2008). In the proximal half of the mandible, the condyle (as part of the condyloid process) is assigned an individual partition to contrast its role as component of the TMJ, distinguished from the other regions, which are the main attachment and loading sites for masticatory muscles (Rinker 1954).
H <sub>5</sub> : {1}{2}{3}{5}{4,6} H <sub>6</sub> : {1,4 <sup>+</sup> }{2}{3}{5}{4 <sup>-</sup> ,6}	Partitions defined to reflect localized functional interactions: {1,4 <sup>+</sup> }/{1}, and {2} span incisor and molar alveoli, whereas {3} and {4,6}/{4 <sup>-</sup> ,6} span attachment sites and loading regions of the temporalis, and lateral masseter and pterygoid muscles, respectively (Rinker 1954); {5} includes the condyle, which is functionally related to the TMJ. The two models are distinguished by the allocation of the ascending portion of the ramus (sampled by partition {4 <sup>+</sup> }) as part of the incisor alveolus (H <sub>5</sub> ) or as a masseter attachment (H <sub>6</sub> ).
H <sub>7</sub> : {1,2}{5}{3,4,6}	Partitions defined to reflect broad functional interactions distinguishing tooth-bearing portions of the rostral half of the mandible from the muscle-bearing proximal half and the articular region.
H <sub>8</sub> : {1,2}{3,4,5,6} H <sub>9</sub> : {1,2,4 <sup>+</sup> }{3,4 <sup>-</sup> ,5,6}	The mandible is divided into tooth- and muscle-bearing parts. The two versions of this model are distinguished according to whether the ascending ramus is interpreted as insertion site of the masseter (H <sub>8</sub> ) or as the posterior portion of the incisor alveolus (H <sub>9</sub> ). This two-part model has been used in quantitative genetics of the mouse mandible (Cheverud et al. 1997; Klingenberg et al. 2004), and in studies of morphological integration (Klingenberg et al. 2003).
H <sub>10</sub> : {1,2,5}{3,4,6} H <sub>11</sub> : {1}{2,4,5}{3,6} H <sub>12</sub> : {1,2,4,5}{3,6}	Partitions predict integration between condyle and dental alveoli due to the effect of rotation at the TMJ on occlusion patterns (Björk 1969; Björk and Skieller 1972). Models differ in whether the ramus is affected, and whether an effect is predicted for both molar and incisor alveoli or for molar alveolus only.

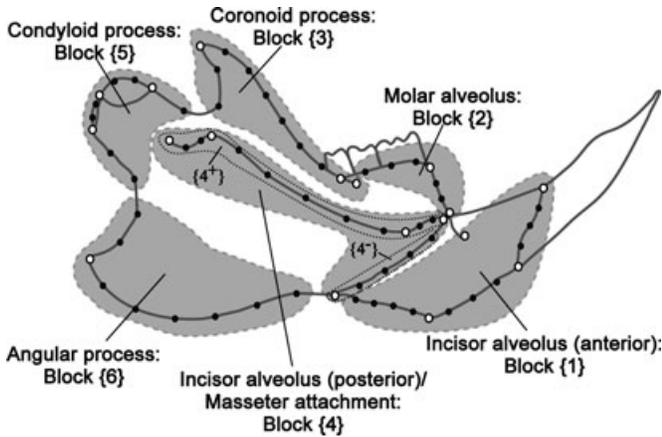
modularity (excluding H<sub>0</sub>) were also tested for support alongside the original models, for a total of 620 different models. Each of these models is thus defined as a unique combination of modules predicted by two or more distinct hypotheses.

The following sections detail the method used to generate and test specific predictions of the models. Briefly, expectations were obtained by allowing coordinates within each module to (co)vary as an independent subspace. Expectations were then compared to the data using a multivariate metric, whose significance was assessed using a Monte Carlo procedure.

#### GENERATION OF EXPECTATIONS FROM DATA

Covariances among observed coordinates are computed in the conventional way (Dryden and Mardia 1998), but expected covari-

ances are obtained by transforming each dataset to match the covariation structure implied by the models. Specifically, Procrustes residuals are partitioned into anatomical regions corresponding to the modules specified by each model, and these partitions are each assigned to their own subspace, orthogonal to all other partitions. That is achieved by the following steps: (1) make as many copies of the original dataset as there are modules in the hypothesis being tested; (2) within each copy, assign a value of zero to each coordinate that does not belong to the module; (3) combine all copies into a single matrix by stacking them vertically, so that the resulting matrix has dimensions  $nm \times 2k$ , where  $n$  is sample size,  $m$  is the number of modules in the model being considered, and  $k$  is the number of (2-D) landmarks and semilandmarks. For example, if we wish to compute an expectation from a model



**Figure 3.** Partitions of the mandible used to define a priori models of modularity.

with three modules containing the coordinates for landmarks 1–8:

$$[1\ 2\ 3] [4\ 5\ 6] [7\ 8],$$

where the numbers in brackets represent the landmarks included in each module, we first partition the full dataset with  $n$  observations and eight variables into three subspaces, by forming the extended data matrix

$$X_0 = \begin{bmatrix} 1 & 2 & 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 5 & 6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 7 & 8 \end{bmatrix}$$

in which each element corresponds either to one of the variables (vectors 1–8) or to a vector of zeros, each of length  $n$ .

The expected covariance matrix, namely  $S_0$ , can be computed from the columns of  $X_0$ . For models of modularity (see above),  $S_0$  will equal the observed covariance matrix after between-module covariances are replaced by zeros. However, to take into account the covariances between Procrustes residuals induced by GLS superimposition (Walker 2000), entries in the extended data matrix ( $X_0$ ) are superimposed prior to computing covariance matrices.

When combining a priori hypotheses to produce 620 mixed models as described above, overlapping modules, that is, modules that share landmarks, are produced. Although module overlap may be biologically reasonable in that two or more processes can affect the same anatomical region, assigning a variable simultaneously to more than one subspace presents a methodological challenge because duplicating a variable in this way will also duplicate its variance and covariances with other variables within the region of the overlap. Therefore, expected covariance matrices must be adjusted so that the model does not excessively depart from observed values. To that end, overlapping modules can be modeled

as contributing equally to the variation of the overlapping region by dividing the Procrustes residuals in matrix  $X_0$  by  $\sqrt{s}$ , where  $s$  equals the number of modules sharing the corresponding landmark. For example, if the model being tested is

$$[1\ 2\ 3] [3\ 4\ 5] [3\ 5\ 6\ 7\ 8]$$

in which variable 3 is shared by three modules, and variable 5 is shared by two,  $X_0$  would equal

$$\begin{bmatrix} 1 & 2 & 3/\sqrt{3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3/\sqrt{3} & 4 & 5/\sqrt{2} & 0 & 0 & 0 \\ 0 & 0 & 3/\sqrt{3} & 0 & 5/\sqrt{2} & 6 & 7 & 8 \end{bmatrix}.$$

It is interesting to note that overlapping modules are no longer orthogonal, allowing the definition and testing of biologically informed patterns that could not possibly be obtained by conventional eigenanalysis techniques.

**ESTIMATING GOODNESS OF FIT**

Goodness of fit between expected and observed patterns can be assessed using a variety of metrics that measure the similarity between covariance matrices such as trace correlations (Klingenberg et al. 2003), Procrustes distances (Peres-Neto and Jackson 2001), matrix correlations (Cheverud 1982; Dietz 1983), Common Principal Component Analyses (Phillips and Arnold 1999; Mezey and Houle 2003), angles between subspaces (Zelditch et al. 2006), and  $\gamma$  (Richtsmeier et al. 2005). The present study uses that last metric ( $\gamma$ ), which is computed as:

$$\gamma = \text{trace} \{ (S - S_0)(S - S_0)^T \}$$

(Richtsmeier et al. 2005), where  $S$  and  $S_0$  are the observed and modeled covariance matrices, respectively, and the  $T$  superscript is the transpose symbol.

The similarity between the data and a model is always affected by the number of fixed parameters; consequently, models having more “zero-covariance” elements will regularly appear to be less similar to the data than those having fewer orthogonal subspaces. To control for this artifact,  $\gamma$  values were regressed on the number of fixed parameters contained in each model (i.e., the number of zeros in their respective covariance matrices). Residuals from the regression were added to the expected value (mean) of  $\gamma$ , yielding  $\gamma^*$ , the test statistic for the evaluation of the models. The smallest values of  $\gamma^*$  correspond to the best-supported models. This procedure is appropriate for  $\gamma$ , which is linearly related to the number of fixed parameters in the model, but simulations demonstrate that more complex methods will be needed when using test statistics that have a nonlinear relationship to the number of fixed parameters (e.g., angles between subspaces).

## SIGNIFICANCE TESTS

To test hypotheses of modularity, the expected distribution of  $\gamma^*$  corresponding to each hypothesis was obtained using a parametric Monte Carlo approach in which the model covariance matrix ( $S_0$ ) and the original sample size ( $n$ ) of each model and species were used to parameterize a Wishart distribution (Krzanowski 2000). This is the distribution of covariance matrices, or, more precisely, sums of squares and cross-products (SSCP matrices) of a multivariate normal population. A random variate generator (Krzanowski 2000), implemented in the function WISHRND of Matlab (The MathWorks 2006) was used to generate 1000 random covariance matrices from this distribution, and  $\gamma^*$  was computed between  $S_0$  and these random covariances, giving a probability value for the hypothesis that this value of  $\gamma^*$  is no larger than that between two matrices produced from the same model. A low ( $<0.05$ )  $P$ -value corresponds to large values of  $\gamma^*$ , indicating a large difference between data and model and thus a poorly fitting model.

Random permutations of the original data can also be used as an alternative to this parametric approach. Those permutations generate a distribution of covariance matrices under the model being tested. In this case, Procrustes residuals from individual specimens are randomly permuted within each module postulated by the model, thus preserving the intramodular while destroying the intermodular covariation structure. As in the Monte Carlo approach, covariances from permuted datasets are used to generate a distribution of  $\gamma^*$  under each model, that can be compared to the original value of  $\gamma^*$ .

## DETERMINING MODEL SUPPORT

Given many models, the procedure outlined above might fail to reject two or more competing hypotheses merely because of inadequate statistical power; in cases, as in this study, when several similar models are simultaneously tested, distinguishing statistically between them demands very large samples. Therefore, a more informative approach is to rank models by the strength of their support (i.e.,  $\gamma^*$ ). Confidence intervals for  $\gamma^*$  were obtained using a jackknife resampling method (Manly 2006) in which a randomly chosen subset of 10% of the specimens were dropped from each sample to produce 500 subsamples, from which 95% confidence intervals were computed as the 2.5 and 92.5 percentiles for each model-data comparison (Klingenberg 1996). Finally, a measure of model support (namely “jackknife support”) was computed by counting the proportion of jackknife samples in which a model ranks first (i.e., has the lowest value of  $\gamma^*$ ).

All analyses were programmed and carried out in Matlab (The MathWorks 2006). Most of the methods described herein have been compiled in the software MINT, which is publicly available for download at <http://www-personal.umich.edu/~emarquez/morph>.

## EVALUATING THE METHOD VIA SIMULATIONS

To assess how well the present approach determines support for a known model, a series of simulations were carried out in which random vectors were sampled from a multivariate normal distribution (Krzanowski 2000) with mean vector zero (i.e., to simulate Procrustes residuals) and covariance matrices derived from the 13 models. These matrices were obtained by forcing a multivariate normal random dataset (with mean vector zero and covariance matrix equal to an identity matrix) into the simple structure described by each model. Simulated data were computed for the minimum (39) and maximum (77) sample sizes used in this study. Then the  $\gamma^*$  values were used to determine whether the best-supported model is the one from which the data were derived when fitted to the original set of 13 models and to the complete set of 620 possible model combinations.

## INTERSPECIFIC COMPARISONS

When comparing species with respect to fit among competing models, the statistics  $\gamma$  and  $\gamma^*$  can be interpreted as the distance between each species' data and the model. However, when a single model fits two or more species equally well, it does not necessarily follow that these species are very close to each other because two objects that are equally distant from a third need not occupy the same position, especially in high-dimensional spaces. To increase the precision with which a species' position is determined, it is useful to have a large number of reference points, i.e. additional models. In this study, this was accomplished by constructing a vector of 620  $\gamma^*$  values (one per model being tested) for each species. This vector has two interpretations. The first is as a set of distances between the observed covariation matrix and 620 patterns with known pattern of modularity. The second is as the coordinates for the data in a “model space” centered on a species' covariance pattern. Because each species' model space is potentially centered at a different position, only the direction of these vectors is comparable across species. In this study, those directions were compared by the correlation coefficients between  $\gamma^*$  vectors for pairs of species, providing an indirect comparative approach of their underlying patterns of integration.

## VISUALIZATION OF MODULARITY PATTERNS

To visualize modularity, a variant of Partial Least Squares (Wold 1966; Rohlf and Corti 2000; Bookstein et al. 2002) was used. PLS is a technique that computes the linear combinations of two blocks of variables (e.g., modules) that maximize the covariance between the sets. Like Principal Component Analysis (PCA), PLS produces sets of orthogonal axes on which each variable set can be projected; in PLS, axes are oriented so that the first vector of one block maximally covaries with the second block; the second and successive vectors are constrained to be orthogonal to every other vector in the basis (Rohlf and Corti 2000). Changes in shape

implied by these vectors are then readily computed using standard geometric morphometrics techniques (Bookstein et al. 2002).

In this study, PLS was used to determine the covariation between each module and the whole mandibular shape, thereby visualizing part-whole covariances. This is preferable to visualizing the variation of isolated parts, for two reasons. First, it offers a more realistic interpretation of a part’s variation by ensuring that it lies along axes of variation actually displayed by the mandible. Second, it allows for visualizing regional variation within the context of the variation of the whole, making it possible to detect patterns of covariation both between and within parts. This property makes the part-whole PLS approach effective as an exploratory technique, and it was used as a heuristic in this study to derive additional models whose fit ( $\gamma^*$ ) was compared to the previously tested 620 models derived a priori.

**BUILDING A CONSENSUS MODEL**

Shared elements of supported models can be combined into a “consensus” hypothesis, i.e. a model containing only modules supported in all or most species. Such a consensus could be obtained for each internal node on the phylogeny, but that phylogenetic approach would be better suited for analyses based on a larger number of species and clades. In the present study, a “horizontal” consensus was built by combining models from all sampled species, using both theoretical and heuristic models. A consensus built in this way should not be interpreted as an interspecific average or overall trend; instead, a consensus includes those modules that are common to all or nearly all species, and thus can potentially be interpreted as a set of phylogenetically conserved modules.

*Results*

**SIMULATIONS**

Analysis of simulated datasets confirms that the methods used herein can detect the correct model of modularity if it is included among tested alternatives (Tables 3 and 4). This is the case for analyses of the 13 original models (Table 2) and also for the 620 models formed by combining them. Because the  $\gamma^*$  values are obtained by regressing model-specific  $\gamma$  values on the number of fixed parameters, results are expected to depend on the number of models being evaluated. As in any least-squares procedure, the larger the number of observations (models, in this case), the more stable the results. Not surprisingly, resolution is improved when all 620 models are simultaneously compared, particularly for model H<sub>0</sub>, which predicts a total absence of integration across the mandible (Table 4).

These simulations show that the methods are typically robust to variation in the number of models being examined, but the signal is sometimes obscured for the most highly modular or weakly integrated models, that is, those that predict weak or no covariances across the mandible. Similarly, the methods are reasonably robust to small samples sizes, with the correct model being supported above the others in all but two cases (i.e., the highly modular H<sub>0</sub> and H<sub>1</sub>, Table 3) even when samples are as small as  $N = 39$ . Jackknife support for the correct model from the analysis of 620 models exceeds 75% for all but one model (H<sub>1</sub>), and exceeds 90% for all but four models (Table 4). Again, the models that predict a larger number of smaller modules tend to produce a weaker signal at  $N = 39$ . As might be expected, the signal is improved when analyses are based on large sample sizes

**Table 3.** Results from simulations, showing the measure of fit ( $\gamma^*$  values) obtained from fitting models H<sub>0</sub>–H<sub>12</sub> (Table 2) to the datasets simulated assuming each of these models (Table 2). Lowest  $\gamma^*$  value (in boldface) indicates the most supported model. Only shown dataset with  $N=39$  simulated individuals.

Simulated model	Model support												
	H <sub>0</sub>	H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	H <sub>4</sub>	H <sub>5</sub>	H <sub>6</sub>	H <sub>7</sub>	H <sub>8</sub>	H <sub>9</sub>	H <sub>10</sub>	H <sub>11</sub>	H <sub>12</sub>
H <sub>0</sub>	0.589	0.589	0.585	0.573	0.577	0.588	0.588	0.576	0.588	0.574	0.580	<b>0.569</b>	0.572
H <sub>1</sub>	0.621	0.463	0.467	0.470	0.501	0.459	0.496	<b>0.456</b>	0.477	0.508	0.458	0.475	0.477
H <sub>2</sub>	0.634	0.472	<b>0.348</b>	0.380	0.439	0.474	0.427	0.481	0.529	0.473	0.509	0.512	0.411
H <sub>3</sub>	0.619	0.525	0.477	<b>0.383</b>	0.497	0.537	0.525	0.509	0.502	0.448	0.536	0.506	0.469
H <sub>4</sub>	0.622	0.515	0.435	0.472	<b>0.412</b>	0.514	0.438	0.495	0.541	0.459	0.515	0.545	0.487
H <sub>5</sub>	0.620	0.468	0.473	0.499	0.483	<b>0.445</b>	0.487	0.452	0.456	0.499	0.465	0.479	0.502
H <sub>6</sub>	0.620	0.512	0.446	0.465	0.434	0.481	<b>0.432</b>	0.505	0.520	0.452	0.510	0.527	0.472
H <sub>7</sub>	0.596	0.521	0.547	0.533	0.497	0.492	0.539	<b>0.404</b>	0.421	0.515	0.420	0.512	0.540
H <sub>8</sub>	0.595	0.520	0.555	0.515	0.495	0.490	0.545	0.403	<b>0.364</b>	0.470	0.433	0.504	0.555
H <sub>9</sub>	0.610	0.547	0.512	0.439	0.461	0.552	0.496	0.506	0.504	<b>0.408</b>	0.529	0.534	0.493
H <sub>10</sub>	0.592	0.527	0.543	0.546	0.517	0.508	0.544	0.445	0.481	0.538	<b>0.413</b>	0.525	0.504
H <sub>11</sub>	0.616	0.518	0.538	0.513	0.514	0.517	0.551	0.494	0.483	0.520	0.511	<b>0.431</b>	0.487
H <sub>12</sub>	0.613	0.532	0.483	0.473	0.500	0.547	0.521	0.526	0.539	0.514	0.514	0.480	<b>0.416</b>

**Table 4.** Results from simulations, showing models supported by at least one jackknife subsample of each simulated dataset, among the 620 models that include all possible module combinations of models H<sub>1</sub>–H<sub>12</sub> (see Table 2). Jackknife support measures the proportion of jackknife subsamples in which each model ranks best. A hundred jackknife runs were used in each case. Only shown dataset with  $N=39$  simulated individuals.

Simulated model	Best ranked models	Jackknife support
H <sub>0</sub>	H <sub>0</sub>	100%
H <sub>1</sub>	{1,2}{3}{5}{4,6}	46%
	{1,2}{3}{4}{5}{6}	42%
	{1}{2}{3}{5}{4,6} (H <sub>6</sub> )	6%
	{1}{2}{3}{4}{5}{6} (H <sub>1</sub> )	3%
	{1}{2}{4}{5}{3,6}	2%
	{1,2}{4}{5}{3,6}	1%
H <sub>2</sub>	{1,4}{2}{3}{5}{6} (H <sub>2</sub> )	94%
	{1,2,4}{3}{5}{6}	6%
H <sub>3</sub>	{1,2,4}{3,5,6} (H <sub>3</sub> )	100%
H <sub>4</sub>	{1,4 <sup>+</sup> }{2}{5}{3,4 <sup>-</sup> ,6} (H <sub>4</sub> )	98%
	{1,4}{2}{5}{3,4 <sup>-</sup> ,6}	2%
H <sub>5</sub>	{1}{2}{3}{5}{4,6} (H <sub>5</sub> )	81%
	{1}{2}{5}{3,4,6}	19%
H <sub>6</sub>	{1,4 <sup>+</sup> }{2}{3}{5}{4 <sup>-</sup> ,6} (H <sub>6</sub> )	75%
	{1,4}{2}{3}{5}{4 <sup>-</sup> ,6}	24%
	{1,4 <sup>+</sup> }{2}{5}{3,4 <sup>-</sup> ,6} (H <sub>4</sub> )	1%
H <sub>7</sub>	{1,2}{5}{3,4,6} (H <sub>7</sub> )	94%
	{1}{2}{5}{3,4,6}	6%
H <sub>8</sub>	{1,2}{3,4,5,6} (H <sub>8</sub> )	78%
	{1}{2}{3,4,5,6}	22%
H <sub>9</sub>	{1,2,4 <sup>+</sup> }{3,4 <sup>-</sup> ,5,6} (H <sub>9</sub> )	99%
	{1,2,4 <sup>+</sup> }{5}{3,4 <sup>-</sup> ,6}	1%
H <sub>10</sub>	{1,2,5}{3,4,6} (H <sub>10</sub> )	100%
H <sub>11</sub>	{1}{2,4,5}{3,6} (H <sub>11</sub> )	100%
H <sub>12</sub>	{1,2,4,5}{3,6} (H <sub>12</sub> )	100%

(data not shown); thus, at  $N = 200$ , jackknife support is 100% even for model H<sub>1</sub>.

Although both  $\gamma^*$  and jackknife support must be interpreted with caution when sample sizes are small, the purpose of these statistics is limited to finding the best-supported hypothesis. However, the pattern of support for the full set of hypotheses, not simply the best one, is a function of the covariation structure in the data, so even though a large sample size may be required when assessing a highly modular model, its support relative to other models will generally be higher when it is true than when it is not. Hence, with  $N = 39$ , H<sub>0</sub> ranks first of 620 when the data are simulated according to H<sub>0</sub>, whereas it ranks between 306<sup>th</sup> and 596<sup>th</sup> when the data are simulated using other models. In contrast, H<sub>1</sub> ranks fifth when data actually follow H<sub>1</sub>, just below models {1,2}{3}{5}{4,6} (first), {1,2}{3}{4}{5}{6},

{1,2}{4}{5}{3,6}, and {1}{2}{3}{5}{4,6} (fourth), but it ranks between 48<sup>th</sup> and 293<sup>rd</sup> when data follow every other model except H<sub>0</sub> and H<sub>5</sub>. In those cases, H<sub>1</sub> ranks second and eighth, respectively. Note that a higher rank for H<sub>1</sub> when data are simulated based on H<sub>0</sub> rather than H<sub>1</sub> does not imply that H<sub>0</sub> is more strongly supported by these data, because ranks are relative measures and regressions are mutually independent. In fact, H<sub>0</sub> ranks 533<sup>rd</sup> when data are simulated using H<sub>1</sub>.

#### FITTING MODULARITY HYPOTHESES TO OBSERVED COVARIATION PATTERNS

Only one model, H<sub>0</sub>, the null model of no integration, is consistently rejected for all species in Monte Carlo tests ( $P < 0.001$ ). That all other models are supported suggests that they are too similar to each other to be discriminated given available sample sizes. Therefore, when evaluating the strength of support for the alternative models, emphasis is placed on relative measures of support (i.e.,  $\gamma^*$  and jackknife support), determined both for the original set of 13 models and the full set of 620 models. Results from both sets, however, are consistent, so that relative ranks of the smaller set are preserved in the larger one, and therefore only results from the latter are discussed.

For all but three species, a single model can be recognized as most strongly supported, having 95% or better jackknife support, meaning that it ranked first out of 620 in at least 95% of 500 jackknife subsamples (Table 5). Somewhat surprisingly, the same hypothesis is supported by all these species. According to this hypothesis, one module encompasses the proximal half of the mandible, spanning the mandibular processes through to the proximal portion of the incisor socket; two additional modules span the rostral portion of the incisor alveolus and the molar alveolus, respectively. In terms of the original 13 models, the first module can be derived from model H<sub>8</sub>, and the other two from H<sub>1</sub>, H<sub>5</sub>, and H<sub>11</sub> (Table 2).

For the other three species, *Microryzomys minutus*, *Oryzomys palustris*, and *Holochilus chacarius*, more than one model seems to fit the data. In the case of the first two species, a single hypothesis is supported in over 50% of the jackknife runs and this model is the same one described above (Table 5). However, four models are supported equally well by the data of the last species, *H. chacarius*. The most common elements of all four models are the module comprising the molar alveolus ({2}, jackknife support = 100%), the condyloid process ({5}, jackknife support = 95%), and the rostral portion of the incisor alveolus ({1}, jackknife support = 73%), plus modules that associate the coronoid with the angular processes ({3,6,\*}, jackknife support = 100%).

#### INTERSPECIFIC COMPARISONS

The same model is the most highly supported by the data from eight of nine species, but there is more differentiation among

**Table 5.** Best-supported models for each species. Given are the percentage of jackknife subsamples in which each model was the best supported, the measure of model fit ( $\gamma^*$ , lowest value highlighted) and its 95% confidence interval, and  $P$ -values for the null hypothesis that observed covariance data belong to the same Wishart distribution as the covariance structure of each model (see text for details). Ranks based on the full set of 620 models that include all possible combinations of models  $H_1$ – $H_{12}$  (see Table 2). Jackknife support and 95% CI based on 500 jackknife subsamples.  $P$ -values based on 1000 Monte Carlo replicates.

Species	Best ranked models	Jackknife support	$\gamma^*$ (95% CI)	$P$
<i>Holochilus chacarius</i>	{1}{2}{5}{3,4,6}	28%	0.250 (0.236–0.261)	1.000
	{1,4 <sup>+</sup> }{2}{5}{3,4 <sup>-</sup> ,6} ( $H_4$ )	27.2%	<b>0.246</b> (0.238–0.256)	1.000
	{1}{2}{4}{5}{3,6}	26.6%	0.247 (0.236–0.261)	0.995
	{1}{2}{4}{5}{3,4 <sup>-</sup> ,6}	13.6%	0.247 (0.237–0.261)	1.000
	{1}{2}{4}{3,5,6}	4.4%	0.254 (0.244–0.270)	1.000
	{1,4 <sup>+</sup> }{2}{3,4 <sup>-</sup> ,5,6}	0.2%	0.261 (0.253–0.270)	1.000
<i>Melanomys caliginosus</i>	{1}{2}{3,4,5,6}	96.4%	<b>0.227</b> (0.207–0.241)	1.000
	{1}{2}{5}{3,4,6}	3.6%	0.239 (0.220–0.260)	1.000
<i>Microryzomys minutus</i>	{1}{2}{3,4,5,6}	63.6%	<b>0.183</b> (0.164–0.193)	1.000
	{1,4}{2}{3}{5}{6} ( $H_2$ )	33%	0.186 (0.176–0.194)	1.000
	{1}{2}{5}{3,4,6}	2%	0.190 (0.169–0.201)	1.000
	{1}{2}{3}{4}{5}{6} ( $H_1$ )	0.8%	0.189 (0.183–0.196)	1.000
	{1}{2}{4}{3,5,6}	0.4%	0.195 (0.185–0.205)	1.000
	{1,4}{2}{3,5,6}	0.2%	0.192 (0.186–0.197)	1.000
<i>Nectomys squamipes</i>	{1}{2}{3,4,5,6}	100%	<b>0.207</b> (0.196–0.218)	1.000
<i>Oligoryzomys nigripes</i>	{1}{2}{3,4,5,6}	99.4%	<b>0.204</b> (0.187–0.218)	1.000
	{1}{2}{5}{3,4,6}	0.6%	0.224 (0.205–0.237)	1.000
<i>Oryzomys couesi</i>	{1}{2}{3,4,5,6}	95.4%	<b>0.190</b> (0.172–0.204)	1.000
	{1}{2}{5}{3,4,6}	4.6%	0.198 (0.176–0.215)	1.000
<i>Oryzomys palustris</i>	{1}{2}{3,4,5,6}	81.8%	<b>0.231</b> (0.213–0.249)	1.000
	{1,4}{2}{3,4 <sup>-</sup> ,5,6}	13%	0.238 (0.229–0.250)	1.000
	{1,4}{2}{3,5,6}	3.6%	0.243 (0.235–0.252)	1.000
	{1}{2}{5}{3,4,6}	0.8%	0.244 (0.226–0.261)	1.000
	{1,4}{2}{5}{3,4,6}	0.4%	0.250 (0.237–0.264)	1.000
	{1,2,4}{3,5,6} ( $H_3$ )	0.2%	0.255 (0.247–0.264)	1.000
	{1,2,4}{3,4 <sup>-</sup> ,5,6}	0.2%	0.250 (0.243–0.263)	1.000
	{1}{2}{3,4,5,6}	98.2%	<b>0.222</b> (0.198–0.240)	1.000
<i>Oryzomys xantheolus</i>	{1,4}{2}{3,5,6}	1.2%	0.246 (0.233–0.260)	1.000
	{1,4}{2}{3,4 <sup>-</sup> ,5,6}	0.6%	0.243 (0.232–0.254)	1.000
<i>Sigmodontomys alfari</i>	{1}{2}{3,4,5,6}	100%	<b>0.189</b> (0.170–0.211)	1.000

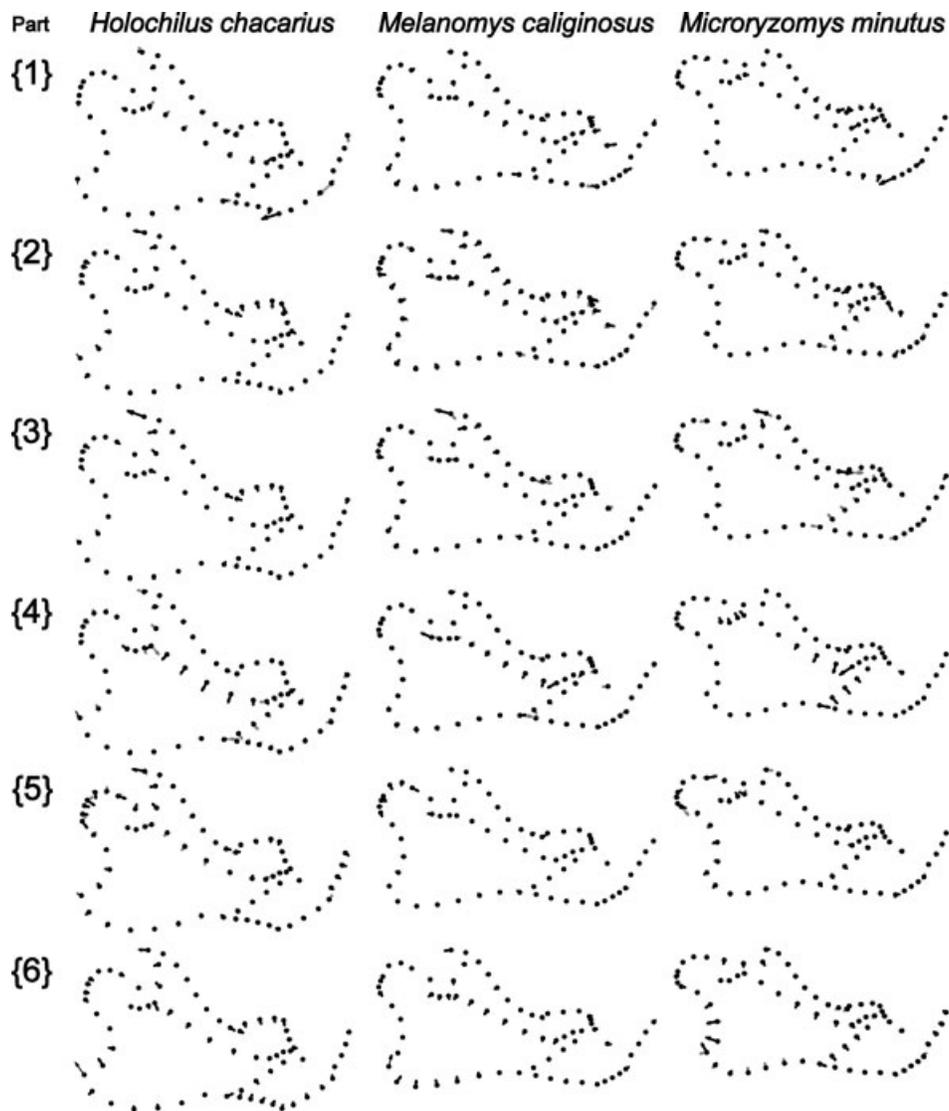
species in structure of modularity than might be inferred from the similar rankings of models. That differentiation is revealed by the correlations between the vectors of 620  $\gamma^*$  values used to compare species in model space (Table 6, Fig. 5). These correlations demonstrate higher similarity among the covariation patterns of *O. xantheolus*, *N. squamipes*, *Melanomys caliginosus*, and *S. alfari*, and of *Oligoryzomys nigripes* and *M. minutus* relative to other species, thus showing a pattern of similarity that coincides with the phylogenetic topology for these species (Fig. 1). Such coincidence is not seen in the three remaining species (*O. couesi*, *O. palustris*, and *H. chacarius*), which also belong to a single clade, but show little similarity to each other in their covariation patterns (Fig. 5).

**PART-WHOLE PARTIAL LEAST SQUARES ANALYSIS**

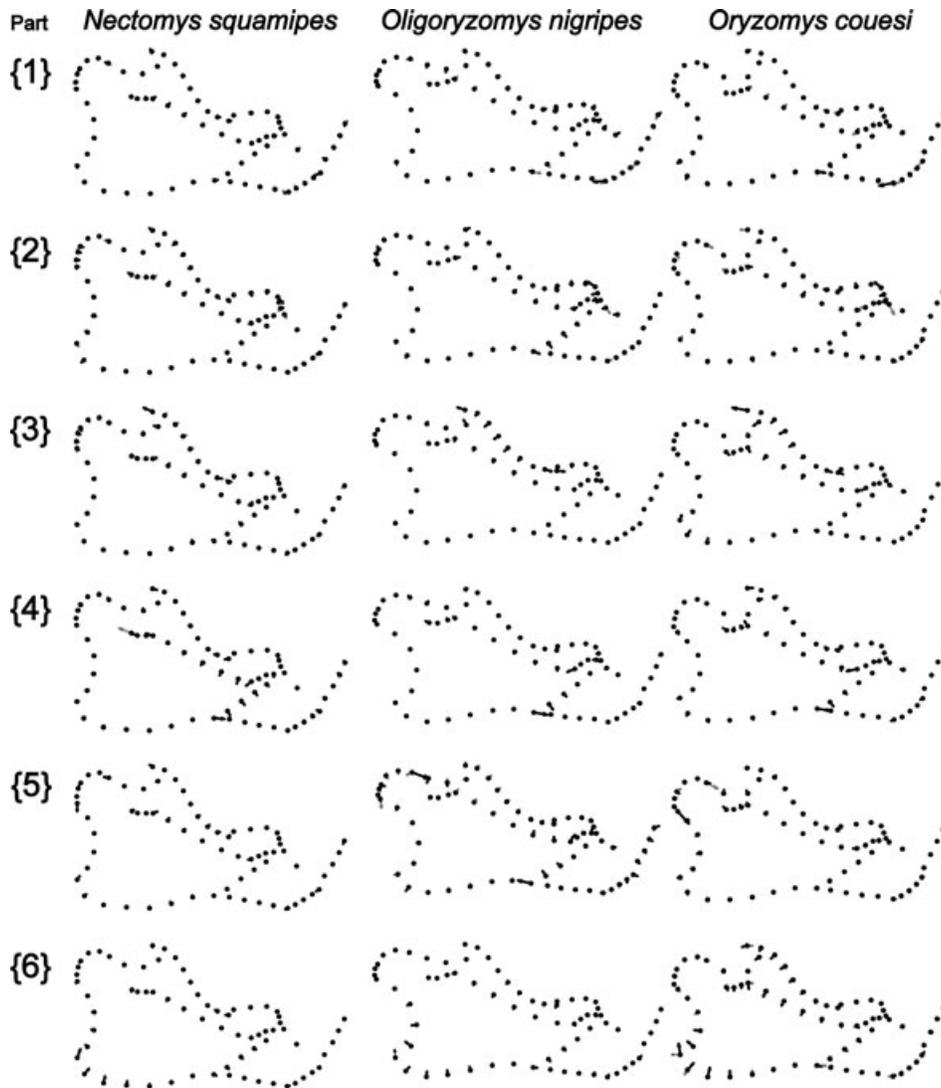
Part-whole regressions examined by PLS analysis are roughly consistent with results from previous tests (Fig. 4). As would be expected, an isolated part covaries with the version of itself embedded in the whole. Other landmarks and regions outside of that partition are not constrained to contribute to the PLS axes and do so only if they covary with landmarks within the fixed partition. For example, when partition {2} in *H. chacarius* is regressed against the whole mandible, a “correlated response” is detected for landmarks belonging to this partition, but there are also covariances with landmarks in other partitions (i.e., {1}, {3}, {4}, and {6}). Similarly, when partition {6} is regressed on the whole mandible, correlated responses are detected

**Table 6.** Pearson's product-moment correlation coefficients among vectors of  $\gamma^*$  values computed for each species for the full set of 620 models (original models in Table 2 plus the combination of their modules). Correlation coefficients measure similarity between species in model space.

	<i>H. chacarius</i>	<i>M. caliginosus</i>	<i>M. minutus</i>	<i>N. squamipes</i>	<i>O. nigripes</i>	<i>O. couesi</i>	<i>O. palustris</i>	<i>O. xantheolus</i>	<i>S. alfari</i>
<i>H. chacarius</i>	1.0								
<i>M. caliginosus</i>	0.948	1.0							
<i>M. minutus</i>	0.718	0.813	1.0						
<i>N. squamipes</i>	0.907	0.968	0.871	1.0					
<i>O. nigripes</i>	0.797	0.906	0.867	0.900	1.0				
<i>O. couesi</i>	0.718	0.849	0.957	0.886	0.909	1.0			
<i>O. palustris</i>	0.859	0.952	0.889	0.955	0.947	0.925	1.0		
<i>O. xantheolus</i>	0.858	0.946	0.787	0.967	0.854	0.832	0.946	1.0	
<i>S. alfari</i>	0.807	0.923	0.851	0.961	0.919	0.911	0.951	0.952	1.0



**Figure 4.** Results from part-whole Partial Least Squares analysis for each species sampled in this study, showing the shape deformations implied by the first (black arrows) and second (gray arrows) PLS vectors after regressing each partition indicated at the left margin onto whole mandible configurations. See text and Figure 3 for partition nomenclature.



**Figure 4. Continued.**

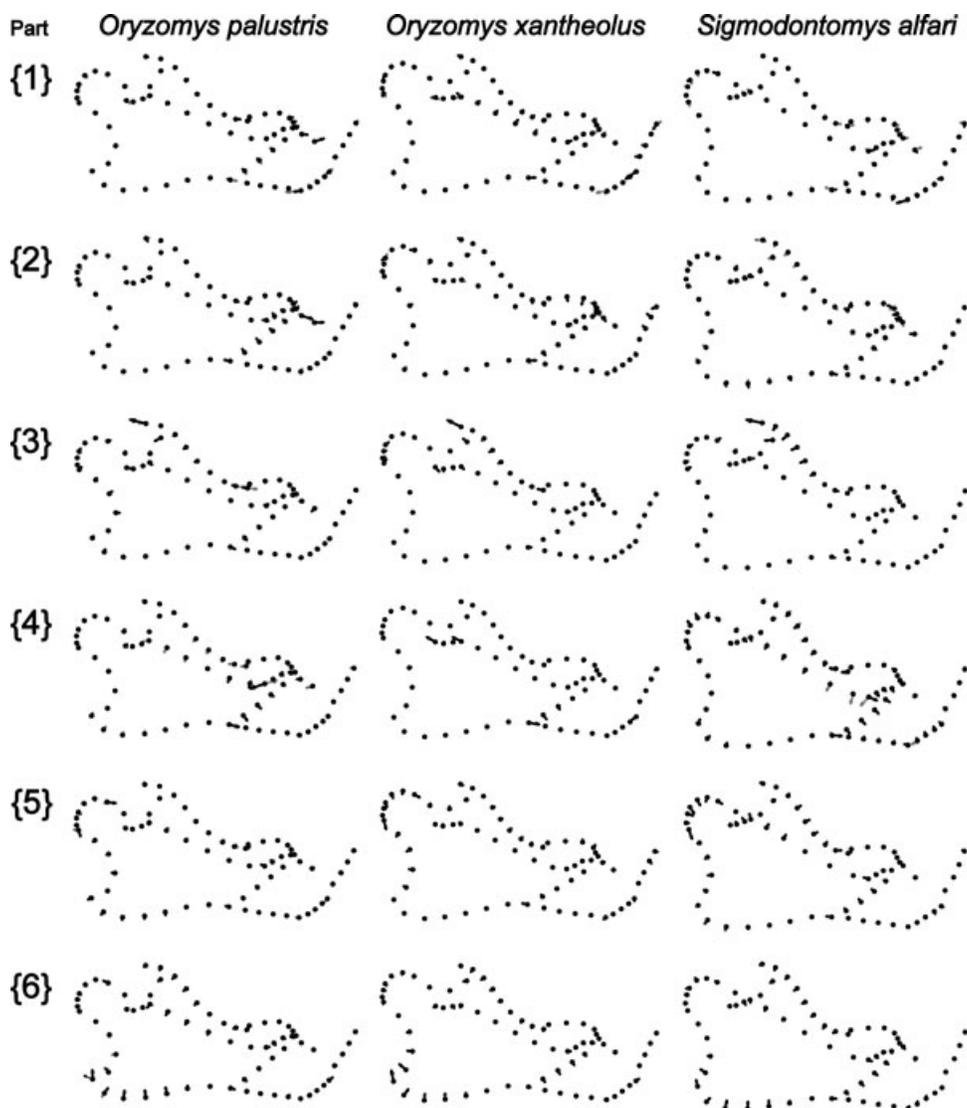
on landmarks belonging to partitions {1}, {2}, {3}, {4}, and {5}.

Comparison of part-whole PLS axes among sampled species reveals even greater diversity in patterns of covariation than detected by comparing positions of species in the model space. For example, two species that would be judged as highly similar in their patterns of modularity based on their patterns of model support are *N. squamipes* and *M. caliginosus* (Fig. 5). There are some important similarities between them that are revealed by PLS, most notably, the association between partitions {3}, {4}, {5}, {6} and a close association between those partitions and partition {1}. At the same time, these species differ in other features, such as the magnitude of the covariation between coronoid ({3}) and angular ({6}) processes, and the specific condylar region ({5}) that covaries with the other partitions, and also the apparent association between rostral edge of the molar alveolus ({2}) and proximal half of the mandible ({3,4,5,6}).

#### HEURISTIC MODELING

Based on visual inspection of patterns revealed by part-whole PLS regression, new models were created, emphasizing patterns of bidirectional covariation between pairs of partitions—that is, when covariation between two partitions is observed irrespective of which one is the “part” in a part-whole analysis. These models were used as primers for a heuristic search; these primers differed among species and the partitions incorporated in these models did not consistently match those deduced a priori from developmental expectations. In the heuristic search, model primers were systematically altered by iteratively adding and removing landmarks and recomputing  $\gamma^*$ , keeping only those changes that lowered its value, until a local minimum was found.

Species-specific models resulting from this search procedure are shown in Figure 6. As shown in Table 7, support for these models is significantly higher than for any of the a priori hypotheses, as might be expected, as the lowest  $\gamma^*$  value for any a priori



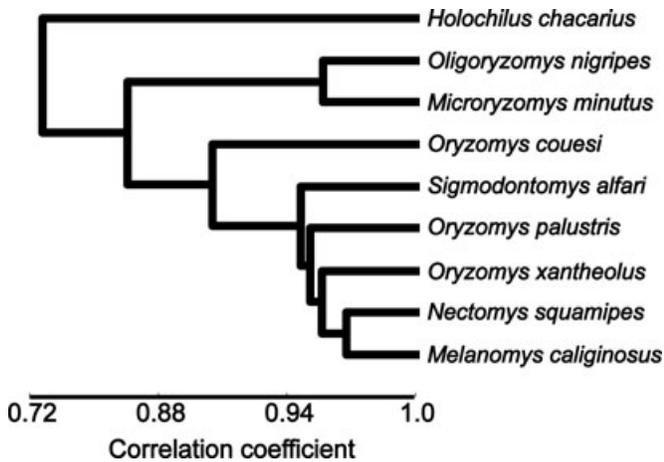
**Figure 4. Continued.**

hypothesis (Table 5) lies outside the 95% confidence interval of the heuristic model in all species. Furthermore, jackknife support is over 98% in all cases. These results suggest that the heuristic models, although harder to interpret biologically than the a priori ones, are nonetheless closer to the position of each species' underlying pattern of modularity within the model space.

Similarities between the heuristic models and the most strongly supported a priori models are mainly restricted to the proximal half of the mandible, comprising sites of muscle attachment and the condyle, parts that often behave as a relatively integrated region in that they covary more highly with each other than with other parts. Exceptions tend to be species-specific, including: (1) the ramus and condyle do not appear to be integrated with the angular and coronoid processes in *H. chacarius*; (2) the coronoid process is more integrated with the ramus than it is with the angular process in *O. nigripes*, and (3) the ramus is integrated

with the condyloid and angular processes, but not the coronoid in *O. palustris* and *S. alfari*. Even though models predicting these associations among mandible regions were among the 620 hypotheses initially tested, their lack of statistical support typically results from differences in the location of boundaries between putative modules. For instance, whereas the a priori models treat the coronoid process as a single module ({3}), heuristic analyses often distinguish its proximal and rostral margins as being separate partitions (Fig. 5).

The most visible differences between heuristic and a priori models concern partitions within the rostral and alveolar regions. In particular, heuristic models tend to group the landmarks of the molar alveolus ({2}) into one or two independent modules, except in three species, in which either the labial or rostral alveolar margin is integrated with other mandible regions. Similarly, heuristic models tend to subdivide the rostral portion of

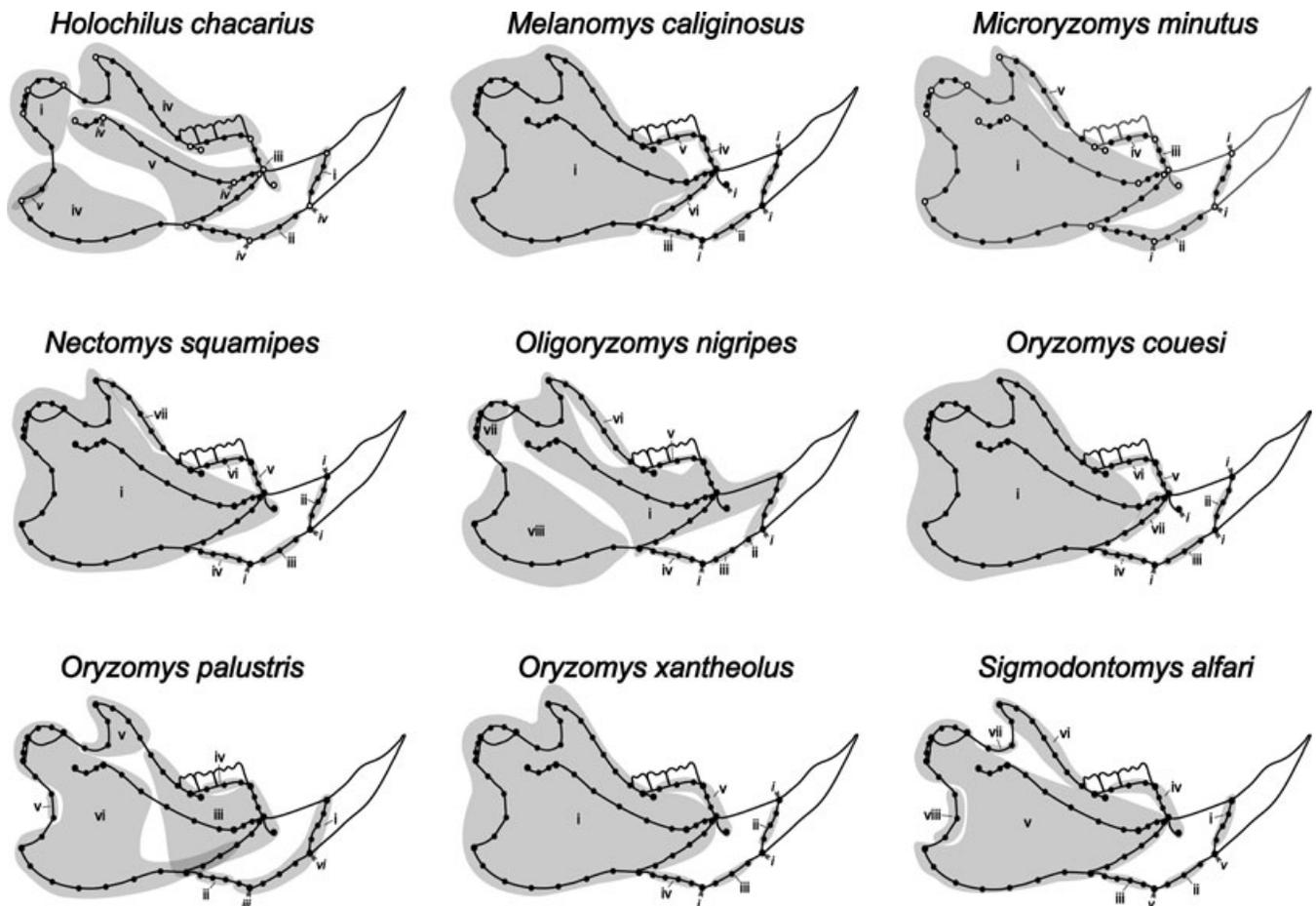


**Figure 5.** UPGMA dendrogram illustrating patterns of interspecific similarities in model space, as measured by correlation coefficients between vectors of model fit to data (i.e.,  $\gamma^*$  values) including all models tested in this study. Cophenetic correlation coefficient = 0.73.

the incisor alveolus ( $\{1\}$ ) into two or more partitions. Subdivisions comprising only semilandmarks are usually recognized as partitions, whereas others, usually comprising landmarks, tend to covary with regions in the proximal half of the mandible (i.e., partitions  $\{4\}$ ,  $\{5\}$ , and  $\{6\}$ ; Fig. 5). This difference between landmarks and semilandmarks seems to be due to a reduced variance of the latter, likely a consequence of their fewer degrees of freedom (Bookstein 1997). It is interesting to note that supported a priori and heuristic models contain a limited degree of module overlap, suggesting either that the processes responsible for within-module covariation do not overlap spatially, that they overlap completely (e.g., modules are hierarchical), or that covariation patterns at overlapping regions are driven by only one of these processes, possibly the most recent one.

#### CONSENSUS MODEL

The consensus derived from the most highly supported a priori hypotheses (Table 5) contains either the partitions  $\{1\}\{2\}\{5\}$



**Figure 6.** Best-supported heuristic models for each of the species sampled in this study. Numbers indicate individual modules (note that some modules are not contiguous). Italicized numbers indicate single landmarks that belong to larger modules elsewhere in the mandible. Overlapping modules or regions are illustrated as dark gray areas. See text for details.

**Table 7.** Best-supported heuristic models for each species. Given are the percentage of jackknife subsamples in which each model was the best supported, the measure of model fit ( $\gamma^*$ , lowest value highlighted) and its 95% confidence interval, and  $P$ -values for the null hypothesis that observed covariance data belong to the same Wishart distribution as the covariance structure of each model (see text for details). Ranks based on the full set of 621 models that include all possible combinations of models  $H_1$ – $H_{12}$  (see Table 2) plus the heuristic model. Jackknife support and 95% CI based on 500 jackknife subsamples.  $P$ -values based on 1000 Monte Carlo replicates.

Species/Model	Jackknife Support	$\gamma^*$ (95% CI)	$P$
<i>Holochilus chacarius</i>	100%	0.173 (0.156–0.193)	1.000
<i>Melanomys caliginosus</i>	100%	0.167 (0.145–0.180)	1.000
<i>Microrhynchomys minutus</i>	100%	0.121 (0.093–0.133)	1.000
<i>Nectomys squamipes</i>	100%	0.159 (0.143–0.169)	1.000
<i>Oligoryzomys nigripes</i>	100%	0.155 (0.130–0.172)	1.000
<i>Oryzomys couesi</i>	100%	0.131 (0.108–0.152)	1.000
<i>Oryzomys palustris</i>	98.4%	0.193 (0.175–0.216)	1.000
<i>Oryzomys xantheolus</i>	100%	0.176 (0.153–0.195)	1.000
<i>Sigmodontomys alfari</i>	99.6%	0.166 (0.145–0.187)	1.000

{3,4,6} or {1}{2}{3,4,5,6} (see Fig. 3). These typically supported modules are the rostral portion of the incisor alveolus ({1}) and the molar alveolus ({2}) and a larger module containing the proximal half of the mandible. The ambiguity in the latter region arises from the independence of the condyloid process ({5}) in *H. chacarius*. However, even when the same theoretical model fits several species, those species can still differ strikingly from each other in model space (Fig. 5). Therefore, a more accurate consensus can be derived from the heuristic models. However, computing a consensus among heuristic models on a landmark-by-landmark basis such as done for theoretical models would yield a large number of partitions composed by only a few landmarks, mainly because differences in boundaries between partitions are generally small, and do not seem sufficient to warrant defining entirely different modules. Thus, certain partitions are considered as equivalent when estimating the consensus model even if they comprise slightly different sets of landmarks.

The consensus hypotheses derived from the heuristic approach differs from that deduced from the theoretical models in at least three respects. First, the module comprising the ramus plus mandibular processes is supported by the data of only five species in heuristic analyses. Data from the remaining four species suggest the presence of four modules: (1) coronoid-angular, (2) ramus-coronoid, (3) condyloid-angular-ramus, and (4) condyloid-angular (Fig. 6). A strict consensus among the models supported by all species would then conclude that each individual region is a separate module, as predicted by model  $H_1$  (Table 2). In fact, aside from discrepancies in the boundaries between regions, the modules detected by heuristic search do not contradict those defined a priori with a few exceptions, such as the unanticipated combination of a part of the ramus plus coronoid process specific to *O. palustris* (Fig. 6). Also, as noted above, the rostral portion of the incisor alveolus seems to have a complex relationship with other

mandibular regions, the only constant element being a statistical association with the part of the ramus that spans the posterior portion of the incisor alveolus (i.e., partition {4}). Finally, the molar alveolus is supported as an independent module in seven of the species, forming associations with the ramus and incisor and with the ramus and coronoid process, respectively, in the remaining two. Nevertheless, the molar alveolus still appears to be a module, or perhaps two, should its rostral and labial margins be considered independent units in the consensus.

These consensual elements can be combined into a single model, denoted as {1,4}{2}{3}{5}{6}. Although this appears to be equivalent to  $H_2$  (Table 2), it is not entirely so because of variation among species in module boundaries, as well as the detection of small modules within some of those partitions, particularly within {1} and {2}, which suggest an increased independence of these modules. Nevertheless, the modules contained in this consensus correspond well to the mesenchyme condensations that give rise to the adult mandible (sensu Atchley and Hall 1991). Being composed by variational modules intrinsically shared among all sampled species, this model is then interpreted as the best approximation to a phylogenetically conserved pattern of integration.

## Discussion

Variational modules are expected to be internally coherent, semi-autonomous units, comprising subsets of traits that covary or correlate highly relative to traits outside these subsets (Olson and Miller 1958; Wagner and Altenberg 1996). Because covariation within modules results from shared processes that span each module as a whole, multiple processes that spatially overlap their effects can blur boundaries between modules, creating covariances between them. Dissecting the complex of processes that

create integration both within and among modules requires analytic methods that can take complex structures of modules and varying levels of trait covariation into account when testing causal hypotheses. The methods outlined in this article are suited to complex, high-dimensional structures because they are based on statistical tests that use all the information contained within a covariance matrix. For that reason, they enable testing models that hypothesize interactions both within and among modules. This feature contrasts the present method with others that treat modules as self-contained units, such as Partial Least Squares (Klingenberg and Zaklan 2000; Bastir and Rosas 2005; Monteiro et al. 2005), approaches that can measure only pairwise associations between modules.

The general approach used in this study rests on the idea that individual modules result as an effect of local processes, which means that the structure of covariation is shaped partly by the accumulation of local effects even when they partially overlap (Cowley and Atchley 1992). Mathematically, the regions in which modules overlap are therefore modeled by assuming that their (co)variances result from the cumulative effect of the overlapping processes. The specific model for those effects used in the present study distributes the variance evenly among the overlapping modules, a choice that is open to improvement. But regardless of how these variances are distributed it is important to take such overlap into account when modeling modules as the dynamic, complex, and partially overlapping units postulated by theory (Hallgrímsson et al. 2007; Wagner et al. 2007). That view of modules shifts the focus of studies of modularity away from asking whether a particular module exists, or whether a given phenotypic part is a module, to asking about the degree to which specific genetic and epigenetic factors contribute to the distinction of modules and, more generally, to the patterning of phenotypic covariances.

### MODULES AS SUBSPACES

Analyzing modularity in terms of both the within- and between-module associations, within the span of full covariance matrices, may seem directly contrary to the idea that modules are defined as independent subsets of variables. However, the present approach reconciles these apparently conflicting views by treating modules as self-contained multidimensional subspaces that are embedded within the full space occupied by the data. This geometric interpretation of modules makes it straightforward to derive and test explicit hypotheses based on developmental and functional theories because the expectations derived from them are merely reorganizations, not modifications of the geometric characteristics (e.g., degrees of freedom) of the original space. In fact, it is this commensurability between data and models, as well as among models, which allows for testing multiple competing hypotheses simultaneously in terms of a unique goodness-of-fit statistic

(Richtsmeier et al. 2005). Furthermore, because all hypotheses are directly compared to the data, results can be continually reexamined by adding new hypotheses or by refining the old ones. An important consequence of this is that there is no penalty for testing novel or even unrealistic hypotheses along with familiar ones. As shown in this study, the fit of a model to the data can be optimized by adjusting the boundaries between modules and allowing further subdivision of putative modules through a heuristic search. The combination of rigorous tests of a priori hypotheses regarding mandibular modularity, plus exploratory optimization of the best-supported models, reveals a substantial diversity in patterns of mandibular modularity among oryzomyine species that could not have been discovered by either approach on its own.

The representation of variational modules as semi-independent subsets of variables (i.e., subspaces) is consistent with their definition as dissociable parts of the phenotype (Raff 1996; Magwene 2001; Mitteroecker and Bookstein 2007). This representation assumes that modules, like most phenotypic features, are inherently multidimensional, which means that individual vectors (e.g., Principal Components) rarely can correspond to an entire module. As a result, methods that extract such vectors are likely to fail to find modules, especially when the modules are not statistically independent. In general, nonindependence (i.e., nonorthogonality) of modules can be due to the spatiotemporal overlap of their causal factors. Therefore, defining modularity in terms of an array of strictly orthogonal subspaces (as done for the models  $H_1$ – $H_{12}$  tested in this study), should be decided in light of developmental and functional considerations rather than methodological limitations (Steppan et al. 2002; Mezey and Houle 2003). We clearly need methods that can detect overlapping modules if the structure of modularity and integration regularly arises from multiple, spatially overlapping processes.

There are other methods that could allow for nonorthogonal modules, such as Confirmatory Factor Analysis (Jöreskog and Wold 1982; Zelditch and Carmichael 1989), but the approach presented herein differs from those alternatives by not relying on assumptions about the expected structure of variation (Fornell and Bookstein 1982; Bookstein 1986). However, there is a methodological limitation that arises from allowing modules to be nonindependent in that nonorthogonal factors (i.e., models or components thereof) cannot be easily combined to compute their joint effects. This limitation can be overcome by extracting axes of maximum covariation among separate modules, using methods such as PLS or canonical correlations (Rohlf and Corti 2000). However, as discussed above, these methods ignore within-module covariances. The alternative presented here, namely part-whole PLS regression, can find the directions of both within- and between-module covariation using the entire covariance matrix. Despite the large number of part-whole comparisons that need to be done (Fig. 4), this is a useful tool for detecting patterns of

intermodule covariation as well as for testing the internal integrity of putative modules.

Simulations show that the methods used herein tend to be robust at small sample sizes, although robustness seems partly dependent on the dimensionality of the data; when there are a large number of modules, sample sizes need to be proportionally increased to facilitate their detection. These simulations, however, represent worst-case scenarios because each of the 52 pairs of coordinates in the simulated samples was treated as a landmark, producing a 100-dimensional space. The oryzomyine data occupy in turn a space with substantially fewer degrees of freedom, as only 19 of the 52 sampled coordinates correspond to landmarks, with the remaining being semilandmarks, each of which contribute only one dimension to the space (Bookstein 1997). Consequently, the sample size requirements should be correspondingly smaller.

Robustness of results depends also on the total number of models being tested, including the original hypotheses, the models that result from combining them, and their heuristic modifications. That dependence arises from the fact that the main test statistic,  $\gamma^*$ , is computed by a least-squares regression (of  $\gamma$  on the number of zero-covariance elements within each model). Each point in that regression represents the value for an individual model. Yet, these results also appear to be highly robust to variation in the number of tested models based on the simulations using 13 versus 620 models (Tables 3 and 5, respectively). The results differ only in their adjustment of the  $\gamma$  value for the model having the greatest number of zero-covariances, that is  $H_0$ , the null model that predicts a total absence of integration. This model behaves as an outlier only when the regression is based on 13 models. When the number of models to be tested is insufficient, outliers such as this one can either be removed or the  $\gamma$  statistic can be adjusted using methods of robust regression.

Finally, results from these tests might be influenced differently by random and nonrandom measurement error. Random (uncorrelated) errors that are large relative to variation among specimens are potentially able to obscure patterns of modularity and can be usually minimized by averaging across replicates of single specimens. If random errors are relatively small, however, they can be generally ignored as they will contribute equally to all comparisons between data and models. Random error is more likely to be a concern when using complicated sampling protocols, such as those that imply setting up specimens in a specific angle prior to measurement (e.g., when analyzing 2D projections of 3D structures). In contrast, large correlated errors may actually produce a covariation structure that could cause random variation to appear modular, or obscure an underlying pattern of modularity by superimposing it. Correcting for nonrandom measurement errors may require estimating their covariation structure and subtracting it from the phenotypic covariance matrix. In addition to systematic errors of any kind, correlated errors are more likely

to be an issue when separate structures (e.g., detached bones) are measured following a different sampling protocol for each.

### COMPARATIVE ANALYSES OF INTRASPECIFIC PATTERNS OF MODULARITY

The simultaneous testing of multiple models not only allows for assessing alternative hypotheses relative to each other, it also produces a useful construct for comparing multiple datasets—a model space, which is the set of dimensions onto which each model can be mapped as a point. This space is useful for measuring differences both among models and between models and data. When only data-model comparisons are of interest, as in this study,  $\gamma^*$  values can be used to map each model in this space by its relative distance (i.e., fit) to the data. Because the same set of models are fitted to all species, it is possible to compare species in terms of their relative distances to each model. Herein these comparisons are done by correlating the data-model vectors of  $\gamma^*$  values, producing a pattern of interspecific differences based on similarities to models.

It should be noted that the questions addressed by interspecific comparisons in model space are different from those addressed by comparing covariance matrices because a model space is defined with respect to hypothesized factors with explicit genetic or epigenetic interpretations. In contrast, direct comparisons between covariance matrices are difficult to interpret in terms of the processes responsible for their differences. That difficulty arises because a unique set of interactions among genetic or epigenetic factors can produce an entire distribution of covariance matrices just by modifying the variance of these factors (Wagner 1984; Zelditch et al. 1990; Boughner and Hallgrímsson 2008). The pattern of interspecific differences found in this study bears a moderate resemblance to the pattern of phylogenetic relationships among oryzomyine species (Fig. 1) in that two of the clades cluster together, although the third one shows no identifiable pattern (Fig. 5). In contrast to this moderate resemblance to phylogenetic history, there is no clear correspondence between the clusters and either geography or habitat distribution (see Supporting Table S1). This could mean that there is a phylogenetic rather than ecological signal underlying the covariance structures of these species.

The search for shared dimensions of variation reveals a set of conserved modules embedded in the covariation structure of most of these species. These subspaces may be interpreted as “intrinsic” variational modules that have persisted despite divergence in both mean phenotypes and covariance patterns. The complete set of persistent modules can then be used to define a consensus model that corresponds to a “latent” pattern of modularity shared by all sampled species. Just like intraspecific modules can be compared to the factors of Confirmatory Factor Analysis, the set of phylogenetically persistent modules within the consensus are analogous to the common factors found by Common Principal

Component Analysis (Phillips and Arnold 1999). The advantage of the present approach is that the persistent modules that it detects are homologous sets of dimensions that have a constant meaning in phenotypic terms, and these dimensions are not forced to be orthogonal. Should persistent modules be mutually orthogonal, as is the case for these oryzomyine mandibles, the persistent modules found by the present approach will be at least conceptually identical to the common factors of CPCA. However, the ability of CPCA to find such modules depends on the internal dimensionality of each module, as CPCA cannot subdivide the space occupied by the data into distinct multidimensional subspaces.

Because persistent modules are a geometric consequence of extracting the shared subspaces from multiple samples, there is no a priori reason to expect that these modules would correspond to actual developmental processes. Nevertheless, the persistent modules detected in the present study do closely map to the mesenchymal condensations that are developmental modules of the mammalian mandible (Atchley and Hall 1991; Hall and Miyake 2000; Hall 2003). Such match strongly suggests that persistent modules result from conserved developmental processes. However, those persistent units do not necessarily correspond to the modules supported by within-species analysis, as shown in this study. Instead, persistent modules detected in the oryzomyine mandibles sampled herein are embedded within subspaces that span larger regions. Those larger subspaces can be regarded as “super-modules” and the persistent modules can be viewed as their basic building blocks, combined in various arrangements in different species. For example, the condylar region, which is highly integrated in all species, is at least partially integrated with the ramus in *M. caliginosus*, *M. minutus*, *N. squamipes*, *O. couesi*, *O. palustris*, *O. xantheolus*, and *S. alfari* (Fig. 6). In most species, intraspecific patterns seem broadly consistent with results obtained from QTL mapping (Mezey et al. 2000) in that the proximal part of the mandible appears to be more tightly integrated than the alveolar region and that they form separate modules. The hierarchical structure of modularity, in which developmental modules nest within super-modules, has been already recognized in other developmental systems (Cowley and Atchley 1992; Beldade et al. 2002; Breuker et al. 2007; Armbruster et al. 2004), and could suggest even deeper hierarchical levels, in which super-modules within species are grouped to form increasingly inclusive clade-level modules. It is worth noting, however, that the observed patterns are not strictly hierarchical (see Fig. 6) in that several modules are not nested within either the proximal or alveolar region of the mandible.

The interpretation of the shared subspaces as phylogenetically conserved modules allows for tracing the origin of modules to early developmental processes. Tracing them to early development assumes that these modules are conserved through ontogeny as well as phylogeny, even if they are embedded within super-

modules. Such an interpretation implies that covariation patterns arising early in development are retained in adults, which would be expected should those early processes be more variable than those occurring later (Hallgrímsson et al. 2007), or not disrupted by processes that occur later. An important question raised by the high degree of conservation of variation patterns ascribed to early development, and the high diversity of those attributed to later processes, is whether the developmental mechanisms responsible for divergence of the latter can be traced to variation of specific factors regulating mandibular morphogenesis and growth. More generally, it should be possible to determine whether phylogenetically conserved modules do in fact act as basic building blocks from which complex morphologies can be patterned by selection. If persistent modules do correspond to those basic blocks, variation among species in patterns of (super-) modularity might arise from divergent functional associations between blocks (Breuker et al. 2006). An apparent developmental constraint, such as the conservative covariance pattern established early in development, may actually contribute to the evolvability of the mandible by enhancing its ability to respond to the diverse ecological demands ultimately responsible for structuring functional integration.

#### ACKNOWLEDGMENTS

Thanks to M. L. Zelditch for her generous assistance in the process of writing this manuscript, to L. L. Knowles and members of her laboratory for their valuable insight in the early stages of designing this study, and to two anonymous reviewers for their helpful comments on an earlier version of this manuscript. Thanks also to P. Myers and S. Hinshaw (UMMZ), M. Carleton, and H. Kafka (USNM), M. Hafner (LSUMZ), and B. Patterson (FMNH) for their generous assistance with obtaining access to their respective specimen collections. This work was funded by NSF Doctoral Dissertation Improvement Grant No. 0407570, a Fellowship of Graduate Student Travel from the Society for Integrative and Comparative Biology, and a Hinsdale Fellowship from the University of Michigan Museum of Zoology.

#### LITERATURE CITED

- Ackermann, R. R., and J. M. Cheverud. 2000. Phenotypic covariance structure in tamarins (genus *Saguinus*): a comparison of variation patterns using matrix correlation and common principal component analysis. *Am. J. Phys. Anthropol.* 111:489–501.
- Albertson, R. C., J. T. Streebman, T. D. Kocher, and P. C. Yelick. 2005. Integration and evolution of the cichlid mandible: the molecular basis of alternate feeding strategies. *Proc. Natl. Acad. Sci. USA* 102:16287–16292.
- Armbruster, S. W., C. Pélabon, T. F. Hansen, and C. P. H. Mulder. 2004. Floral integration, modularity, and accuracy: distinguishing complex adaptations from genetic constraints. Pp. 23–49 in M. Pigliucci and K. Preston, eds. *Phenotypic integration: studying the ecology and evolution of complex phenotypes*. Oxford Univ. Press, Oxford.
- Atchley, W. R., and B. K. Hall. 1991. A model for development and evolution of complex morphological structures. *Biol. Rev. Camb. Philos. Soc.* 66:101–157.
- Bastir, M., and A. Rosas. 2005. Hierarchical nature of morphological integration and modularity in the human posterior face. *Am. J. Phys. Anthropol.* 128:26–34.

- Beldade, P., K. Koops, and P. M. Brakefield. 2002. Modularity, individuality, and evo-devo in butterfly wings. *Proc. Natl. Acad. Sci. USA* 99:14262–14267.
- Berg, R. L. 1960. The ecological significance of correlation pleiades. *Evolution* 14:171–180.
- Björk, A. 1969. Prediction of mandibular growth rotation. *Am. J. Orthodont.* 55:585–599.
- Björk, A., and V. Skieller. 1972. Facial development and tooth eruption: an implant study at the age of puberty. *Am. J. Orthodont.* 62:339–383.
- Bookstein, F. L. 1986. The elements of latent variable models: a cautionary lecture. Pp. 203–229 in M. E. Lamb, A. L. Brown, and B. Rogoff, eds. *Advances in developmental psychology*, vol. 4. Lawrence Erlbaum, Hillsdale, NJ.
- . 1991. *Morphometric tools for landmark data*. Cambridge Univ. Press, New York.
- . 1997. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Med. Image Anal.* 1:225–243.
- Bookstein, F. L., A. P. Streissguth, P. D. Sampson, P. D. Connor, and H. M. Barr. 2002. Corpus callosum shape and neuropsychological deficits in adult males with heavy fetal alcohol exposure. *Neuroimage* 15:233–251.
- Boughner, J. C., and B. Hallgrímsson. 2008. Biological spacetime and the temporal integration of functional modules: a case study of dento-gnathic developmental timing. *Dev. Dyn.* 237:1–17.
- Breuker, C. J., V. Debat, and C. P. Klingenberg. 2006. Functional evo-devo. *Trends in Ecol. Evol.* 21:488–492.
- Breuker, C. J., M. Gibbs, H. Van Dyck, P. M. Brakefield, C. P. Klingenberg, and S. Van Dongen. 2007. Integration of wings and their eyespots in the speckled wood butterfly *Pararge aegeria*. *J. Exp. Zool. B Mol. Dev. Evol.* 308:454–463.
- Carleton, M. D. 1973. A survey of gross stomach morphology in new-world cricetine rodentia muroidea with comments on functional interpretations. *Misc. Publ. Mus. Zool. Univ. Mich.* 146:1–43.
- Cheverud, J. M. 1982. Phenotypic, genetic, and environmental morphological integration in the cranium. *Evolution* 36:499–516.
- . 2004. Modular pleiotropic effects of quantitative trait loci on morphological traits. Pp. 132–153 in G. Schlosser and G. P. Wagner, eds. *Modularity in development and evolution*. Univ. Chicago Press, Chicago.
- Cheverud, J. M., E. J. Routman, and D. J. Irschick. 1997. Pleiotropic effects of individual gene loci on mandibular morphology. *Evolution* 51:2006–2016.
- Cowley, D. E., and W. R. Atchley. 1992. Quantitative genetic models for development, epigenetic selection, and phenotypic evolution. *Evolution* 46:495–518.
- Dietz, E. J. 1983. Permutation tests for association between two distance matrices. *Syst. Zool.* 32:21–26.
- Draghi, J., and G. R. Wagner. 2008. Evolution of evolvability in a developmental model. *Evolution* 62:301–315.
- Dryden, I. L., and K. V. Mardia. 1998. *Statistical shape analysis*. Wiley, New York.
- 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. *J. Exp. Zool. B Mol. Dev. Evol.* 296:58–79.
- Fornell, C., and F. L. Bookstein. 1982. Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *J. Market. Res.* 19:440–452.
- Guabloche, A., M. Arana, and O. E. Ramirez. 2002. Diet and gross gastric morphology of *Oryzomys xantholus* (Sigmodontinae, Rodentia) in a Peruvian loma. *Mammalia* 66:405–411.
- Hall, B. K. 2003. Unlocking the black box between genotype and phenotype: cell condensations as morphogenetic (modular) units. *Biol. Phil.* 18:219–247.
- Hall, B. K., and T. Miyake. 2000. All for one and one for all: condensations and the initiation of skeletal development. *Bioessays* 22:138–147.
- Hallgrímsson, B., D. E. Lieberman, N. M. Young, T. Parsons, and S. Wat. 2007. Evolution of covariance in the mammalian skull. *Novartis Found. Symp.* 284:164–190.
- Hansen, T. F. 2006. The evolution of genetic architecture. *Annu. Rev. Ecol. Evol. Syst.* 37:123–157.
- Herring, S. W. 1993. Epigenetic and functional influences on skull growth. Pp. 153–206 in J. Hanken and B. K. Hall, eds. *The Skull*, vol. 1: development. Univ. Chicago Press, Chicago.
- Hershkovitz, P. 1944. A systematic review of the Neotropical water rats of the genus *Nectomys* (Cricetinae). *Misc. Pub. Mus. Zool. Univ. Mich.* 58:1–107.
- Hine, E., and M. W. Blows. 2006. Determining the effective dimensionality of the genetic variance-covariance matrix. *Genetics* 173:1135–1144.
- Jöreskog, K. G., and H. Wold. 1982. The ML and PLS techniques for modeling with latent variables: historical and comparative aspects. Pp. 263–270 in H. Wold and K. Jöreskog, eds. *Systems under indirect observation: causality, structure, prediction*, vol. I. North-Holland, Amsterdam.
- Kaufman, M. H., and J. Bard. 1999. *The anatomical basis of mouse development*. Academic Press, San Diego.
- Klingenberg, C. P. 1996. Multivariate allometry. Pp. 23–49 in L. F. Marcus, M. Corti, A. Loy, G. J. P. Naylor, and D. E. Slice, eds. *Advances in morphometrics*. Plenum Press, New York.
- . 2005. Developmental constraints, modules, and evolvability. Pp. 219–247 in B. Hallgrímsson and B. K. Hall, eds. *Variation: a central concept in biology*. Elsevier, Amsterdam.
- Klingenberg, C. P., and S. D. Zaklan. 2000. Morphological integration between developmental compartments in the *Drosophila* wing. *Evolution* 54:1273–1285.
- Klingenberg, C. P., and M. Zimmermann. 1992. Static, ontogenic, and evolutionary allometry: a multivariate comparison in nine species of waterstriders. *Am. Nat.* 140:601–620.
- 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.
- Klingenberg, C. P., K. Mebus, and J. C. Auffray. 2003. Developmental integration in a complex morphological structure: how distinct are the modules in the mouse mandible? *Evol. Dev.* 5:522–531.
- Krzyszowski, W. J. 2000. *Principles of multivariate analysis: a user's perspective*. Oxford Univ. Press, Oxford.
- 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. *Am. Nat.* 153:201–214.
- Magwene, P. M. 2001. New tools for studying integration and modularity. *Evolution* 55:1734–1745.
- Manly, B. F. J. 2006. *Randomization, bootstrap and Monte Carlo methods in biology*, 3rd ed. Chapman and Hall/CRC Press, Boca Raton, FL.
- Márquez, E. J., and L. L. Knowles. 2007. Correlated evolution of multivariate traits: detecting co-divergence across multiple dimensions. *J. Evol. Biol.* 20:2334–2348.
- 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.
- Marshall, L. G., S. D. Webb, J. J. Sepkoski, and D. M. Raup. 1982. Mammalian evolution and the great American interchange. *Science* 215:1351–1357.

- Mezey, J. G., and D. Houle. 2003. Comparing G matrices: are common principal components informative? *Genetics* 165:411–425.
- . 2005. The dimensionality of genetic variation for wing shape in *Drosophila melanogaster*. *Evolution* 59:1027–1038.
- Mezey, J. G., J. M. Cheverud, and G. P. Wagner. 2000. Is the genotype-phenotype map modular?: a statistical approach using mouse quantitative trait loci data. *Genetics* 156:305–311.
- Mitteroecker, P., and F. Bookstein. 2007. The conceptual and statistical relationship between modularity and morphological integration. *Syst. Biol.* 56:818–836.
- . 2008. The evolutionary role of modularity and integration in the hominoid cranium. *Evolution* 62:943–958.
- Monteiro, L. R., V. Bonato, and S. F. dos Reis. 2005. Evolutionary integration and morphological diversification in complex morphological structures: mandible shape divergence in spiny rats (Rodentia, Echimyidae). *Evol. Dev.* 7:429–439.
- Olson, E. C., and R. L. Miller. 1958. Morphological integration. Univ. Chicago Press, Chicago.
- Orr, H. A. 2000. Adaptation and the cost of complexity. *Evolution* 54:13–20.
- Pavlicev, M., J. P. Kenney-Hunt, E. A. Norgard, C. C. Roseman, J. B. Wolf, and J. M. Cheverud. 2008. Genetic variation in pleiotropy: differential epistasis as a source of variation in the allometric relationship between long bone lengths and body weight. *Evolution* 62:199–213.
- Peres-Neto, P. R., and D. A. Jackson. 2001. How well do multivariate data sets match? The advantages of a Procrustes superimposition approach over the Mantel test. *Oecologia* 129:169–178.
- Phillips, P. C., and S. J. Arnold. 1999. Hierarchical comparison of genetic variance-covariance matrices. I. Using the Flury hierarchy. *Evolution* 53:1506–1515.
- Raff, R. 1996. The shape of life: genes, development, and the evolution of animal form. Univ. Chicago Press, Chicago.
- Riedl, R. 1977. Systems-analytical approach to macro-evolutionary phenomena. *Q. Rev. Biol.* 52:351–370.
- Richtsmeier, J. T., S. R. Lele, and T. M. Cole, III. 2005. Landmark morphometrics and the analysis of variation. Pp. 49–69 in B. Hallgrímsson and B. K. Hall, eds. *Variation: a central concept in biology*. Elsevier, Amsterdam.
- Rinker, G. C. 1954. The comparative myology of the mammalian genera *Sigmodon*, *Oryzomys*, *Neotoma*, and *Peromyscus* (Cricetinae), with remarks on their intergeneric relationships. *Misc. Pub. Mus. Zool. Univ. Mich.* 83:1–165.
- Riska, B. 1986. Some models for development, growth, and morphometric correlation. *Evolution* 40:1303–1311.
- Rohlf, F. J. 2006. TpsDig2 software, ver. 2.10. State Univ. New York, Stony Brook.
- Rohlf, F. J., and D. Slice. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Syst. Biol.* 39:40–59.
- Rohlf, F. J., and M. Corti. 2000. Use of two-block partial least-squares to study covariation in shape. *Syst. Biol.* 49:740–753.
- Sheets, H. D. 2002a. Semiland6 software. Canisius College, Buffalo, NY.
- . 2002b. Standard6 software. Canisius College, Buffalo, NY.
- Smith, K. K. 1993. The form of the feeding apparatus in terrestrial vertebrates: studies of adaptation and constraint. Pp. 150–196 in J. Hanken and B. K. Hall, eds. *The Skull*, vol. 3: functional and evolutionary mechanisms. Univ. Chicago Press, Chicago.
- Steppan, S. J. 1997. Phylogenetic analysis of phenotypic covariance structure I. Contrasting results from matrix correlation and Common Principal Component Analyses. *Evolution* 51:571–586.
- Steppan, S. J., P. C. Phillips, and D. Houle. 2002. Comparative quantitative genetics: evolution of the G matrix. *Trends Ecol. Evol.* 17:320–327.
- The MathWorks, Inc. 2006. Matlab R2006a software. Natick, MA.
- Tokita, M., T. Kiyoshi, and K. N. Armstrong. 2007. Evolution of craniofacial novelty in parrots through developmental modularity and heterochrony. *Evol. Dev.* 9:590–601.
- Voss, R. S. 1988. Systematics and ecology of Ichthyomyine rodents (Muroidea): patterns of morphological evolution in a small adaptive radiation. *Bull. Am. Mus. Nat. Hist.* 188:259–493.
- Wagner, G. P. 1984. On the eigenvalue distribution of genetic and phenotypic dispersion matrices: evidence for a nonrandom organization of quantitative character variation. *J. Math. Biol.* 21:77–95.
- . 1988. The influence of variation and of developmental constraints on the rate of multivariate phenotypic evolution. *J. Evol. Biol.* 1:45–66.
- Wagner, G. P., and L. Altenberg. 1996. 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. Pp. 338–358 in G. Schlosser and G. P. Wagner, eds. *Modularity in development and evolution*. Univ. Chicago Press, Chicago.
- Wagner, G. P., M. Pavlicev, and J. M. Cheverud. 2007. The road to modularity. *Nat. Rev. Genet.* 8:921–931.
- Walker, J. A. 2000. Ability of geometric morphometric methods to estimate a known covariance matrix. *Syst. Biol.* 49:686–696.
- Wekslar, M. 2003. Phylogeny of Neotropical oryzomyine rodents (Muridae: Sigmodontinae) based on the nuclear IRBP exon. *Mol. Phyl. Evol.* 29:331–349.
- . 2006. Phylogenetic relationships of oryzomyine rodents (Muroidea: Sigmodontinae): separate and combined analyses of morphological and molecular data. *Bull. Am. Mus. Nat. Hist.* 296:1–149.
- Wold, H. 1966. Nonlinear estimation by iterative least square procedures. Pp. 411–444 in F. N. David, ed. *Research papers in statistics*. Wiley, New York.
- Wright, S. 1932. General, group and special size factors. *Genetics* 17:0603–0619.
- Young, R. L., and A. V. Badyaev. 2006. Evolutionary persistence of phenotypic integration: influence of developmental and functional relationships on complex trait evolution. *Evolution* 60:1291–1299.
- Zelditch, M. L., and A. C. Carmichael. 1989. Ontogenetic variation in patterns of developmental and functional integration in skulls of *Sigmodon fulviventer*. *Evolution* 43:814–824.
- Zelditch, M. L., J. Mezey, H. D. Sheets, B. L. Lundrygan, and T. Garland. 2006. Developmental regulation of skull morphology II: ontogenetic dynamics of covariance. *Evol. Dev.* 8:46–60.
- Zelditch, M. L., D. O. Straney, D. L. Swiderski, and A. C. Carmichael. 1990. Variation in developmental constraints in *Sigmodon*. *Evolution* 44:1738–1747.

Associate Editor: C. Goodnight

## *Supporting Information*

The following supporting information is available for this article:

**Table S1.** Detailed list of specimens sampled for this study, indicating localities of origin and sampling date.

**Table S2.** Results of permutation tests of homogeneity of covariance matrices between sexes and, where applicable, geographical locations and collection dates, for the nine oryzomyine species sampled for this study.

This material is available as part of the online article from:

<http://www.blackwell-synergy.com/doi/abs/10.1111/j.1558-5646.2008.00476.x>

(This link will take you to the article abstract).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting informations supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.