Evolution, 55(9), 2001, pp. 1734–1745

NEW TOOLS FOR STUDYING INTEGRATION AND MODULARITY

PAUL M. MAGWENE
Department of Ecology and Evolutionary Biology, Yale University, Osborn Memorial Laboratories, P.O. Box 208106, New Haven, Connecticut 06520-8106
E-mail: paul.magwene@yale.edu

Abstract.—The study of phenotypic integration concerns the modular nature of organismal phenotypes. The concept provides a rationale for why certain subsets of phenotypic traits show particularly high levels of association over development and/or evolution. The techniques detailed in this report facilitate the generation and testing of hypotheses of phenotypic integration and trait interaction. The approach advocated for exploring patterns of interaction among traits is based on the statistical notion of conditional independence, incorporated in a technique known as graphical modeling. The use of graphical models is illustrated with an application to a well-known biological dataset of fowl skeletal measurements, previously analyzed by Sewall Wright. A definition of phenotypic modularity is given, based on a notion of mutual information, which provides a consistent criterion for recognizing and delimiting integrated subsets of traits and which can be related to standard models of multivariate selection.

Key words.—Correlation, integration, modularity, multivariate selection, phenotypic evolution.

Received November 1, 2000. Accepted May 15, 2001.

> . . . a full and quantitative characterization of modularity would have to allow for hierarchies, gradations, and overlapping of modules. The development of a quantitative characterization of modularity is part of the research program advocated here. (Wagner and Altenberg 1996, pp. 971–972).

Olson and Miller (1951, 1958) developed their concept of morphological integration to explain why, over ontogenetic or evolutionary histories, certain subsets of phenotypic traits should covary together strongly, while other subsets show only weak association. Such patterns of differential association were hypothesized to be the result of underlying developmental or functional causes. Olson and Miller reasoned that uncovering such integrated subsets and exploring the causes and consequences of integration would lead to interesting insights about phenotypic evolution. Unfortunately, the approach that Olson and Miller advocated for studying morphological integration was complex and somewhat obtuse, and the concept initially received only limited attention in the literature, despite the fact that other workers were pursuing research programs with a similar emphasis on patterns of covariation (e.g., Berg 1960; Sokal 1962; Van Valen 1965).

Cheverud and coauthors (Cheverud 1982, 1984; Cheverud et al. 1983) revived interest in the concept of morphological integration by suggesting a quantitative genetic framework (Lande 1979; Cheverud 1982, 1984) for considering the evolution of integration and by providing statistically sound and readily comprehensible tests for hypotheses of morphological integration (Cheverud 1982; Cheverud et al. 1983, 1989). More recently, new information and models about genetic architecture (e.g., Cheverud et al. 1997) and the organization of developmental systems (e.g., Nemeschkal 1999; Rice 2000) has also fostered renewed interest in the topic. Chernoff and Magwene (1999) review some of the recent literature on the concept and discuss in greater detail various methodological approaches to the study of morphological integration.

Various analytical techniques have been advocated for the study of integration. These approaches include such multivariate statistical techniques as cluster analysis (Van Valen 1965; Cheverud 1982), matrix similarity tests (Cheverud et al. 1989), confirmatory factor analysis (Zelditch 1987, 1988), and principal component and standard factor analysis (Cane 1993).

From a statistical perspective, methods that employ statistical tests of a priori hypotheses are preferable; however, it is often the case that prior knowledge of specific biological systems is limited. Empirical evidence, theory, or intuition may provide information about broad patterns of interaction among traits without providing specific hypotheses to be tested. Even if a reasonable model can be formulated and is found to fit relatively well, caution in interpretation is still required given that other untested models may be equally parsimonious. The complexity of possible trait interactions suggests useful roles for both exploratory and confirmatory tools. Even when a specific hypothesis of integration is supported by statistical modeling, there is much to be gained by data exploration and consideration of alternative hypotheses. Tools that can incorporate both types of approaches are likely to be the most useful.

Although various analytical approaches have been advocated for characterizing phenotypic integration, previous discussions have not led to satisfactory methodological definitions of integration or explicit criteria for recognizing integrated subsets of traits. Taking, for the moment, correlation to be the preferred measure of association, what amount of correlation among traits is indicative of integration? The problem of defining appropriate levels of significant interaction was one of the reasons for the rather confusing integrative patterns presented by Olson and Miller (1951, 1958). It is often the case that morphological traits measured on continuous scales exhibit relatively high levels of correlation. This is generally attributed to common patterns of positive association due to growth (Wright 1932; Bookstein et al. 1985), and various maneuvers (Bookstein et al. 1985) have been developed to correct for size in some manner. Even after size is corrected for, one is still faced with the issue of

which subsets of traits should be considered as integrated. A further complicating issue is that even highly integrated traits must have significant interactions outside of their particular subsets if organisms are integrated wholes (Olson and Miller 1958). This suggests that a useful definition of modularity must allow for overlapping and hierarchical patterns of interaction (Wagner and Altenberg 1996).

Biologically speaking, any useful definition of modularity should have explicit links to the larger body of evolutionary theory. The nature of this relationship will depend on the kinds of data being considered and the level of biological organization at which the questions are being addressed. In the context of population phenotypic data, connections to standard mathematical models of multivariate evolution (e.g., Lande and Arnold 1983) are most appropriate.

The issues raised above are tackled in the following pages, and I present a combination of techniques that address analytical requirements for exploring and testing phenotypic datasets for patterns of integration. I begin with a discussion of conditional independence and the role it plays in a family of statistical approaches known as graphical modeling. The utility of conditional independence relationships and graphical modeling for exploring population data is illustrated by presenting an analysis based on correlations among fowl skeletal measurements (Wright 1932; Marcus 1990). The discussion of conditional independence and graphical models serves as a foundation for a new, statistically based, definition of integrated subsets of traits (i.e., phenotypic modules). This concept of modularity is related to standard models for multivariate selection.

Statistical Background

Conditional Independence

The concepts of independence and conditional independence play a critical role in probability theory and statistical inference (Dawid 1979). A thorough discussion of these concepts is beyond the scope of this paper, however, for the reader interested in pursuing a deeper understanding of these concepts, Stirzaker (1994) provides a very readable introduction to probability theory and Whittaker (1990) gives a clear and concise discussion of the probabilistic aspects of independence and conditional independence as they apply to graphical modeling. I attempt to highlight those aspects of the concepts that are particularly relevant to the task at hand—elucidating patterns of biological trait interaction.

Independence encapsulates the notion that pairs of events or distributions of random variates have no systematic relationship to one another. Conditional independence, in contrast, is a statement that two events or random variates are statistically independent of each other, *after* accounting for a third (which may be a set of events or variables). If A and B are independent, conditional on C, the following holds:

$$P(A \cap B \mid C) = P(A \mid C)\, P(B \mid C) \qquad (1)$$

(read: ''The joint probability of the events A and B, conditional on the event C, is equal to the product of the independent probabilities of A and B conditional on C.'')

A convenient notation for this relationship is to write $A \perp B \mid C$. This statement can be read, ''A is independent of B conditional on C.'' Note that conditional independence statements are symmetric (i.e., $A \perp B \mid C$ implies $B \perp A \mid C$; Dawid 1979; Whittaker 1990).

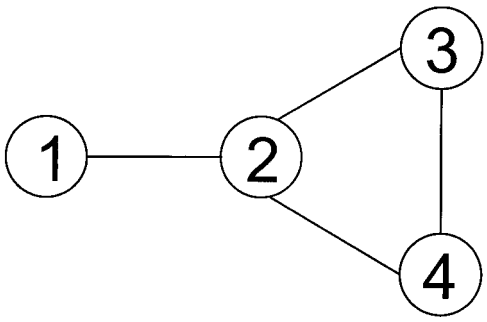

FIG. 1. A graph composed of four vertices and four edges. See text for further explanation.

Lauritzen (1996) provides an appealing, nonprobabilistic framework for considering conditional independence, likening the variables in a conditional independence statement to units of knowledge, such as books. If one thinks of the variables of a conditional independence statement as subjects of study and/or books about those subjects, the expression $X \perp Y \mid Z$ could be read as, ''If one knows Z, studying (a book about) Y is irrelevant for understanding X (and vice versa).''

Independence graphs

A useful way to summarize a series of statements about conditional independence among a set of variables is to use the symbolic language of graph theory.

A graph is a mathematical object composed of two sets: vertices, {V}, and edges, {E}. The set of edges consists of pairs of elements from {V}. It is common to denote the vertices with positive integers: $V = \{1, 2, \ldots, p\}$. Edges can be either directed or undirected. A directed edge implies an ordering of vertices (i, j), whereas an undirected edge, {i, j}, implies that E contains both (i, j) and (j, i). This report is concerned primarily with undirected edges.

One can diagram a graph by drawing vertices as circles and edges as lines. For example, the graph $G = V\{1, 2, 3, 4\}; E\{\{1, 2\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\}$, can be depicted with the diagram in Figure 1.

Two vertices, V_i and V_j, which form an edge are called ''adjacent'' (written $V_i \sim V_j$), or are said to be ''neighbors.'' A path is a sequence of vertices, $V_1, V_2, \ldots, V_m$, for which $\{V_i, V_{i+1}\}$ is an edge in E for all $i = 1, 2, \ldots, m - 1$. Any two vertices, V_i and V_j, are connected if there is a path from V_i to V_j and from V_j to V_i. A subset of vertices is complete if all pairs of vertices in the subset are adjacent. A subset is contiguous if there is a path between all pairs of vertices in the subset. If U is a subset of the vertices in V, the boundary of U is all other vertices in V which are adjacent to a vertex in U (Whittaker 1990; Edwards 1995; Lauritzen 1996).

Graphs are a natural way to summarize patterns of conditional independence among a set of variables, **X**. A graph that summarizes statements about conditional independence is called an independence graph (or a conditional independence graph). An independence graph is an undirected graph,

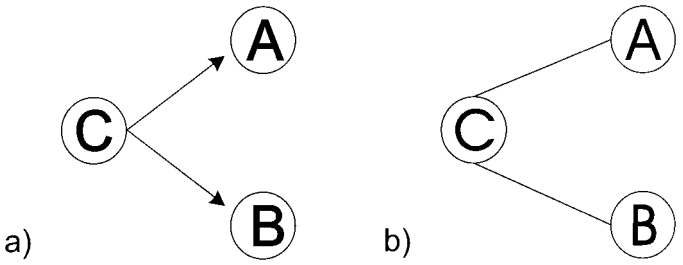

FIG. 2. Two alternative representations of conditional independence. (a) A causal model of independence; (b) a noncausal model of independence.

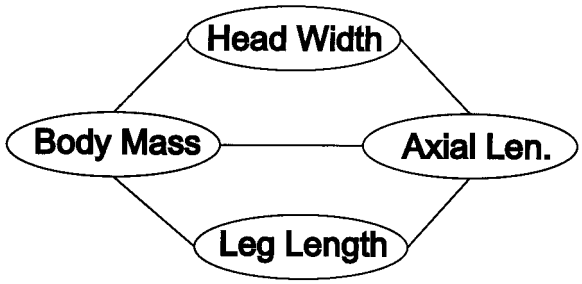

FIG. 3. A hypothetical example of a biological hypothesis depicted as an independence graph. See text for further explanation.

with the variables of $\mathbf{X}$ represented by vertices, in which {i, j} is *not* an edge if and only if $\mathbf{X}_i$ and $\mathbf{X}_j$ are conditionally independent given the rest of the variables (i.e., $\mathbf{X}_i \perp \mathbf{X}_j \mid \mathbf{X}_{K\setminus\{i,j\}}$) (Whittaker 1990). Stated another way, any two variables in an independence graph that are not adjacent are conditionally independent. This property of independence graphs is called the ''pairwise Markov property'' (Edwards 1995).

Interpreting Figure 1 as an independence graph implies the following statements:

$$\mathbf{X}_1 \perp \mathbf{X}_4 \mid (\mathbf{X}_2, \mathbf{X}_3) \quad \text{and} \tag{2a}$$

$$\mathbf{X}_1 \perp \mathbf{X}_3 \mid (\mathbf{X}_2, \mathbf{X}_4). \tag{2b}$$

An even stronger interpretation is available by applying the separation theorem (Whittaker 1990). Let $\mathbf{X}_a$, $\mathbf{X}_b$, and $\mathbf{X}_c$ be nonoverlapping (disjoint) subsets of $\mathbf{X}$. $\mathbf{X}_c$ is said to separate $\mathbf{X}_a$ and $\mathbf{X}_b$ if the independence graph of $\mathbf{X}$ indicates that every path from a vertex in $\mathbf{X}_a$ must go through a vertex in $\mathbf{X}_c$ to get to an element of $\mathbf{X}_b$. Separation implies $\mathbf{X}_a \perp \mathbf{X}_b \mid \mathbf{X}_c$. Applying the separation theorem to the graph in Figure 1, the above statements can be rewritten as:

$$\mathbf{X}_1 \perp \mathbf{X}_4 \mid \mathbf{X}_2 \quad \text{and} \tag{3a}$$

$$\mathbf{X}_1 \perp \mathbf{X}_3 \mid \mathbf{X}_2. \tag{3b}$$

The property implied by the separation theorem is known as the ''global Markov property'' (Whittaker 1990; Edwards 1995).

Another property of independence graphs is the local Markov property (Lauritzen 1996). The local Markov property states that if $\mathbf{X}_b$ is the boundary of $\mathbf{X}_a$, and $\mathbf{X}_c$ is the subset containing all other variables not in $\mathbf{X}_a$ or $\mathbf{X}_b$, then $\mathbf{X}_a \perp \mathbf{X}_c \mid \mathbf{X}_b$. The three Markov properties of independence graphs are equivalent (Whittaker 1990; Lauritzen 1996).

The Role of Conditional Independence in Statistical Theory

Whittaker (1990) argues that conditional independence provides a unifying theoretical framework that ties together the diverse body of methods that fall under the rubric of multivariate statistics. Many techniques, such as path analysis, latent variable modeling, and multivariate regression, are based heavily on the notion of conditional independence (Krzanowksi and Marriott 1995), even if the models are not always stated in such terms. These techniques typically rely on prior information or hypotheses about cause-and-effect relationships among variables, either observed or unobserved, to formulate statements of conditional independence (Krzanowksi and Marriott 1995). For example, one way to read the path diagram in Figure 2a is that variables A and B are independent, conditional on the causal variable C. There is a hypothesis of causality, as indicated by the directed lines from C to A and B. In the absence of prior information about cause and effect among the variates, it is still possible to explore patterns of conditional independence. Figure 2b implies a similar hypothesis of conditional independence in terms of an independence graph. Again, this is a statement that variables A and B are independent conditional on C, but specific causal relationships have not been specified. Graphical models may indeed include causal relationships (Edwards 1995; Lauritzen 1996; Cox and Wermuth 1996), but the use of graphical modeling techniques in this report is limited to undirected independence graphs.

An example of biological interactions that might be formulated in the context of conditional independence is ''Over ontogeny, head width is correlated with leg length, but only as a result of a common association with body mass and the length of the axial column.'' This statement could be reformulated in terms of conditional independence as follows: head width $\perp$ leg length $\mid$ {body mass, axial length}. An independence graph for this statement is given in Figure 3. Body mass and axial length are neither cause nor effect of head width and leg length. One might argue that in this instance body mass and axial length are proxies for some general growth factor, but that interpretation is not necessary. The model simply asserts that given information about body mass and axial length, no further knowledge about head width is gained by knowing leg length and vice versa.

Conditional Independence in Multivariate Normal Distributions

Given p variables, $\mathbf{X} = \{x_1, x_2, \ldots, x_p\}$, which have a multivariate normal (MVN) distribution, the following three statements are equivalent (Krzanowksi and Marriott 1995): (1) the variables x_i and x_j are independent, conditional on $\mathbf{X}_{\{K\}}$ where $\mathbf{X}_{\{K\}}$ stands for any subset of $\mathbf{X}$ that does not include x_i and x_j; (2) The partial correlation, $\rho_{ij\cdot\{K\}} = 0$; and (3) if Σ is the covariance matrix, the element ω_{ij} of the inverse covariance matrix, $\mathbf{\Omega} = \Sigma^{-1}$, is zero. The techniques of graphical modeling, as they apply to continuous variables, use the three equivalence statements listed above as the basis for exploring patterns of interaction among variables.

What about application to non-MVN distributions? Cox

TABLE 1. Correlation matrix, fowl dataset (Marcus 1990). L, skull length; B, skull breadth; H, humerus length; U, ulna length; F, femur length; T, tibia length.

	L	B	H	U	F	T
L	1					
B	0.583	1				
H	0.621	0.584	1			
U	0.603	0.526	0.937	1		
F	0.569	0.515	0.877	0.878	1	
T	0.602	0.548	0.874	0.894	0.926	1

TABLE 2. Inverse correlation matrix, fowl dataset. Variables are defined in Table 1.

	L	B	H	U	F	T
L	1.890					
B	−0.619	1.769				
H	−0.352	−0.996	10.021			
U	−0.261	0.600	−6.916	10.417		
F	0.138	0.219	−2.243	−0.393	8.197	
T	−0.386	−0.465	0.259	−3.076	−5.482	9.087

and Wermuth (1996) note that many of the relevant properties of graphical models follow not from the multivariate normal distribution in particular, but rather from circumstances in which properties of the linear least-squares predictor are relevant. The MVN distribution represents one set of circumstances in which the linear least-squares predictor has a strong justification (Cox and Wermuth 1996). Despite this, much of the material presented here (e.g., significance tests; maximum-likelihood estimates [MLEs]) is only strictly applicable given the assumption of joint normality among the variables.

GRAPHICAL MODELING: AN EXAMPLE

Graphical modeling is a multivariate statistical technique that is primarily concerned with describing and analyzing the interrelationships among observed variables. The potentially large number of interactions among variates is reduced and simplified by conditioning or controlling for other factors using the notion of conditional independence. Graphs are used to provide a visual summary of hypothesized relationships. The techniques may be applied to both continuous and discrete variables (Whittaker 1990; Edwards 1995).

Graphical modeling has its roots in a variety of related areas including path analysis (Wright 1918, 1921, 1932, 1934), statistical physics, and log-linear and covariance selection models (Dempster 1972). Edwards (1995) and Whittaker (1990) both point to a paper by Darroch et al. (1980) as one of the seminal works that combined the various concepts drawn from other approaches in a useful manner.

When describing graphical modeling, it is useful to present a worked example that demonstrates the steps of an analysis. In doing so, I emulate Whittaker (1990), who takes the reader through just such an example in the introductory chapter of his book on graphical models. In the context of this example I will also introduce some of the statistics utilized to assess hypotheses of conditional independence. For a fuller account of the concepts illustrated in this section, I refer the reader to Whittaker (1990), Edwards (1995), and Lauritzen (1996).

Facilities for analysis

Most of the steps involved in the analyses described below can be carried out using any of the common mathematical or statistical software packages (e.g., S-Plus, Matlab). The routines for calculating MLE correlation matrices, model deviances, and a variety of other useful statistics were implemented in the computer programming language Python (for more details, see http://www.python.org). Computer code for these routines is available from the author upon request. Edwards (1995) provides a specialized package for graphical modeling which, among other features, includes facilities for dealing with discrete variables and mixed models.

The Fowl Dataset

One of Sewall Wright's oft-used datasets for demonstrating his method of path analysis was a sample of skeletal measurements based on 276 adult female cross-bred chickens (Wright 1932, 1934, 1954, 1968). The traits Wright considered in his examples were: length of the skull (L), breadth of the skull (B), length of the humerus (H), length of the ulna (U), length of the femur (F), and length of the tibia (T). Marcus (1990) recalculated the correlations between these variables from the original data and arrived at slightly different values than those published by Wright, which he attributed to differences in round-off rules. I use Marcus's (1990) published values to help illustrate the concepts of graphical modeling.

Inverse Covariance

Given a set of variables whose distribution is multivariate normal, no information is lost by condensing the set of variables to a sample mean vector and the sample covariance matrix (Whittaker 1990; Krzanowksi and Marriott 1995). Alternately, one may summarize the same information with a correlation matrix (scaled covariances) and vectors of sample means and standard deviations. The correlation matrix for the fowl dataset, **S**, as given by Marcus (1990) is presented in Table 1.

A less commonly encountered multivariate construct is the inverse covariance matrix, $\Omega = \Sigma^{-1}$. The inverse covariance matrix is also referred to as the ''concentration'' or ''precision'' matrix (Cox and Wermuth 1996). As noted previously, if an element in the inverse covariance (or correlation matrix) is zero, than the corresponding variables are conditionally independent given the remaining variables. The inverse correlation matrix for the fowl data set, $\mathbf{S}^{-1}$, is given in Table 2.

Each diagonal element in the inverse correlation matrix is related to the multiple correlation coefficient between that variable and all the rest; Ω_{ii} equals $1/(1 - R^2)$, where R is the multiple correlation coefficient between variable i and all other variables (Whittaker 1990). The square of the multiple correlation coefficient is known as the ''coefficient of multiple determination'' (Sokal and Rohlf 1995), and, in the context of multiple regression, is a measure of the proportion of variance in variable i jointly explained by all the other variables. For example, the proportion of explained variation

TABLE 3. Partial correlation matrix, fowl dataset. Variables are defined in Table 1.

	L	B	H	U	F	T
L	1					
B	0.338	1				
H	0.081	0.237	1			
U	0.059	−0.140	0.677	1		
F	−0.035	−0.057	0.248	0.042	1	
T	0.093	0.116	−0.027	0.316	0.635	1

TABLE 4. Conditional independence matrix, fowl dataset. Variables are defined in Table 1.

	L	B	H	U	F	T
L	*					
B	*	*				
H	0	*	*			
U	0	*	*	*		
F	0	0	*	0	*	
T	0	0	0	*	*	*

for the humerus variable is R^2(H; rest) = (10.021 − 1)/10.021 = 90.0%, whereas for skull breadth R^2(B; rest) = (1.769 − 1)/1.769 = 43.5% (see Table 2). Humerus length is highly predictable given all the other variables, whereas skull breadth is only mildly so.

If the inverse covariance matrix is scaled to have ones on the diagonal in the same way as a correlation matrix is derived from a covariance matrix, then the off-diagonal elements of the resulting matrix are the negatives of the partial correlation coefficients given all other variables:

$$\rho_{ij\cdot\{K\}} = \frac{-\Omega_{ij}}{(\Omega_{ii}\Omega_{jj})^{0.5}}. \quad (4)$$

The partial correlation matrix for the fowl dataset is listed in Table 3. One notes, for example, that the correlation between femur and tibia length is still strong after conditioning on all other variables, whereas the adjusted correlation between humerus and tibia is nearly zero. Near-zero elements (see below) in the partial correlation matrix suggest that the corresponding traits are conditionally independent.

Conditional Independence Matrices and Graphs

A way to simplify the pattern apparent in the partial correlation matrix (Table 3) is to replace near-zero elements with zeros and replace all other elements with asterisks (Table 4). The issue of how close to zero an element must be is taken up below.

This matrix of zeros and asterisks can be viewed as a hypothesis about patterns of conditional independence for the variables under study. An alternate and more readily interpretable representation of the same information is an independence graph. The independence graph implied by the fowl data is illustrated in Figure 4a. All pairs of variables with nonzero elements in the conditional independence matrix (Table 4) are drawn as adjacent (i.e., they share an edge) in

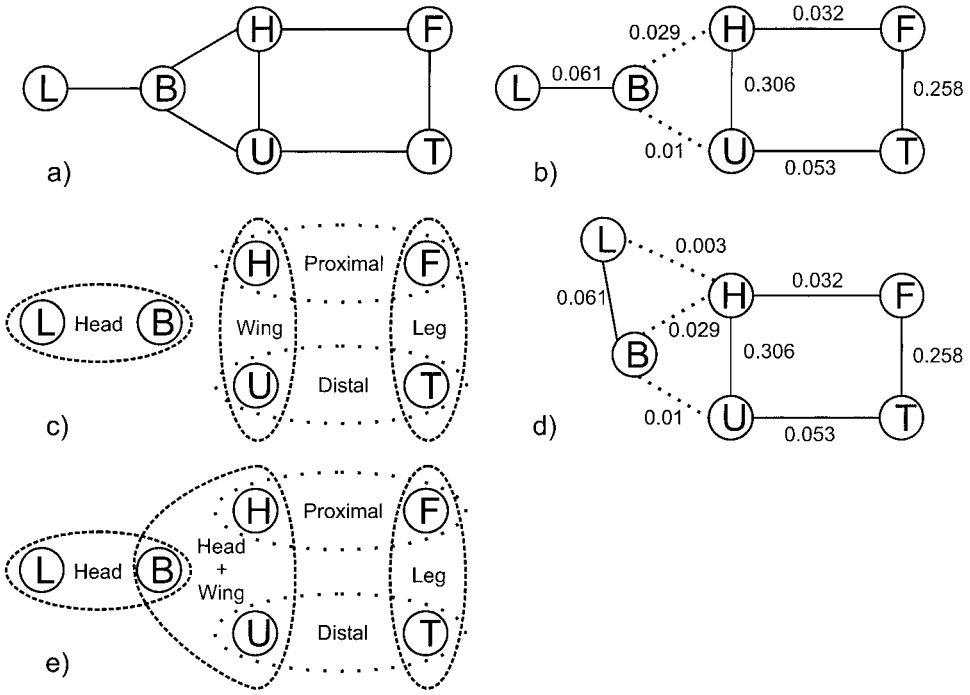

FIG. 4. (a) Independence graph suggested by the inverse correlation matrix for the fowl dataset; (b) edge strengths associated with the independence graph for the fowl dataset; (c) a representation of the compartmentalization of fowl skeletal traits based on the conditional independence model; (d) a modified model of interaction for the fowl dataset that includes an additional edge between head and limb elements; (e) integrated modules according to the strong definition of modularity presented in the text. See text for further explanation. Variables are defined in Table 1.

TABLE 5. Edge exclusion deviance matrix, fowl dataset. Variables are defined in Table 1.

	L	B	H	U	F	T
L	*					
B	33.587	*				
H	1.807	15.903	*			
U	0.956	5.438	169.141	*		
F	0.342	0.91	17.45	0.499	*	
T	2.409	3.73	0.204	29.073	142.574	*

the independence graph. There is no distinction made between negative and positive partial correlations.

Edge Exclusion Deviance

One can test that any one of the specified zero elements is sufficiently close to zero using a statistic called the ''edge exclusion deviance'' (Whittaker 1990). The edge exclusion deviance is an information theoretic measure of whether a particular edge can be eliminated from a saturated model (the model with all possible connection between variables) and is based on the concept of information divergence (Whittaker 1990).

The formula for the edge exclusion deviance is:

$$-N \ln(1 - \rho^2_{ij\cdot\{K\}}), \tag{5}$$

where N is the number of specimens in the sample and $\rho_{ij\cdot\{K\}}$ is the partial correlation coefficient between variables i and j with all other elements held constant (Whittaker 1990; Edwards 1995). A matrix of edge exclusion deviances for the fowl dataset is presented in Table 5.

Each edge exclusion deviance value tests a single edge, not all edges simultaneously. Deviance has an asymptotic χ^2-distribution. The value of each edge is tested against the χ^2-distribution with one degree of freedom, and all edges with deviances less than 3.84 (the 5% point on the χ^2-distribution with df = 1) are rejected (Whittaker 1990). This test supports the pattern of adjacency inferred above, but also points to a borderline case, the edge {B, T}. The model is tentatively accepted as is; in a later section, I will take up the subject of testing the fit of models as a whole to the sample covariance or correlation matrix.

Edge Strength

Information divergence can provide another useful bit of information about the model—the strength of the edges. The strength of an edge {i, j} in a conditional independence graph is measured by the information in $\mathbf{X}_i$ about $\mathbf{X}_j$ and vice versa, conditional on all the rest of the variables. This is expressed: $\mathrm{Inf}(\mathbf{X}_i \perp \mathbf{X}_j \mid \mathbf{X}_{K\setminus\{i,j\}})$ or $\mathrm{Inf}(i \perp j \mid K\setminus\{i, j\})$ (Whittaker 1990). This statistic can be applied to sets of variables as well as to single pairs. For any pair of variables:

$$\mathrm{Inf}(\mathbf{X}_i \perp \mathbf{X}_j \mid \mathbf{X}_K) = -0.5\cdot\ln(1 - \rho^2_{ij\cdot\{K\}}). \tag{6}$$

The edge strength has a lower bound of zero when the partial correlation between $\mathbf{X}_i$ and $\mathbf{X}_j$ is zero.

Applying this measure of strength to the example results in the pattern of edge strengths illustrated in Figure 4b. As stated previously, all the edges drawn are significant (as indicated by their edge exclusion deviances), but the edge strengths indicate that some of the edges are relatively weak, in particular the edges {B, U} and {B, H}. Consequently, in this second figure these edges are drawn as dotted rather than solid lines.

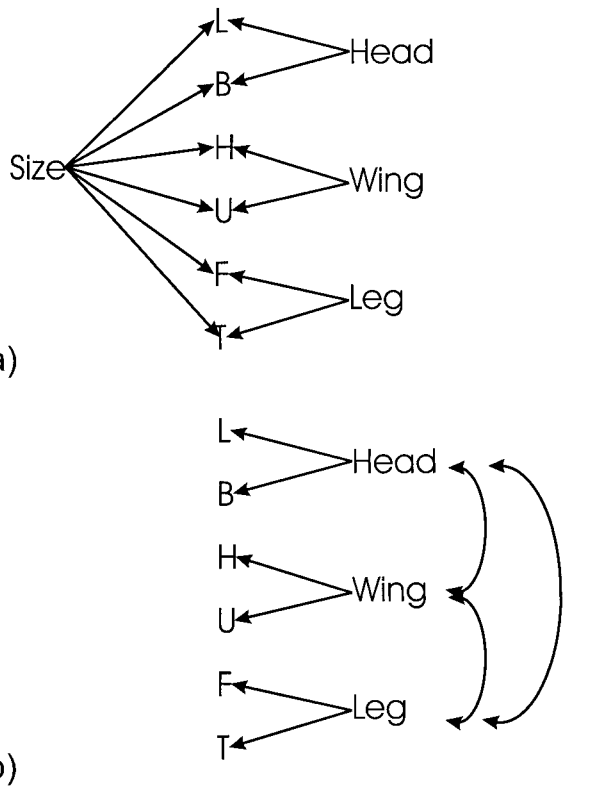

FIG. 5. Two potential path models based on analysis of the fowl dataset. (a) Model suggested by Wright (1932); (b) an alternative model presented by Marcus (1990). Variables are defined in Table 1.

A consideration of the information theoretic underpinnings of edge strength is useful. One way to interpret information is as a measure of the average predictability of one variable from a second (Whittaker 1990). In this sense, one can interpret the edge strengths as a measure of how predictable one variable is from another, with all others held constant. High edge strengths are an indicator of strong copredictability among traits. I return to this point later in my discussion of testing patterns of morphological integration.

Summarizing the above results, one sees that an examination of the inverse covariance or the partial correlation matrix suggests a pattern(s) of conditional independence that can be expressed in an independence graph. The graph depicted in Figure 4a,b is one of a number of possible models that describe the pattern of interactions among the traits under consideration. I return to the question of how well the model fits the data in the next section.

Turning to the issue of interpretation, it is useful to compare the above results to path analyses of the same dataset. Wright (1932) and Marcus (1990) present two possible path models (Fig. 5). The first (Wright 1932) hypothesizes that there are four conditionally independent latent variables responsible for the pattern of correlations—a single general size factor and three factors related to particular compart-

ments of the body, head (L and B), wing (H and U), and leg (F and T; Fig. 5a). This model fits the data relatively well. Marcus (1990) presents an alternative model (Fig. 5b), which includes only the three latent variables—head, wing, and leg—but hypothesizes paths between the latent variables. This model has a similar χ^2 goodness-of-fit statistic, and requires fewer parameters. Marcus (1990) presents a variety of other models that have similarly good fit to the data.

My interpretation of the graphical model presented above (Fig. 4b) agrees in large part with both path models. Examination of the patterns of conditional independence and edge strengths suggests a similar pattern of compartmentalization between head, wing, and leg characters. The interaction between wing and leg elements appears to be much stronger than that between the head elements. There are weaker interactions between head and wing and wing and leg compartments. The interactions between head and wing appear to be the weakest. One might argue based on the graphical model that the implied interaction between head and leg should be deleted from the path model (Fig. 5b; but recall that the edge {B, T} had a borderline edge exclusion deviance).

The graphical model points to an additional set of interactions between fore- and hindlimb elements that are topologically equivalent, that is {humerus, femur}, and {ulna, tibia}, but *not* {humerus, tibia} or {femur, ulna}. This suggests that the pattern of interaction between the limbs may reflect yet another compartmentalization of proximal and distal limb elements that is somewhat weaker than those within forelimb and hindlimb (Fig. 4c). This interaction was not noted in Wright's later works (1954, 1968), nor in recent discussions of path analysis applied to this dataset (Bookstein et al. 1985; Marcus 1990). The current author had not considered such an interaction prior to an analysis using graphical models. Interestingly, one of Wright's early summaries of the method of path analysis hints at a similar interpretation (Wright 1934).

As with the use of path models, it is important to consider alternative hypotheses. The method, as presented so far, has been used in a purely exploratory fashion. A simple examination of the scaled inverse correlation matrix (or partial correlation matrix) was enough to suggest an interpretable pattern of conditional independence among the traits and suggested patterns of interaction that were not immediately obvious from the original patterns of correlation or from the path models previously posited.

Testing Conditional Independence Models

In addition to exploring patterns of conditional independence as suggested by an inverse covariance or partial correlation matrix, graphical modeling techniques can be used to test specific hypotheses. The basic procedure for testing graphical models involves the following steps: (1) specify a hypothesized pattern of conditional independence; (2) generate a MLE of the covariance matrix based on the constraints specified in the hypothesis; and (3) check the goodness-of-fit of the estimated covariance matrix to the sample covariance matrix.

Specifying Models

Specifying a conditional independence model is straightforward, and hypotheses may be expressed as matrices, graphs, or lists of included or excluded edges. The model must simply specify which variates are conditionally independent. No restriction is placed on the values of the nonzero elements.

Estimating Covariance Matrices

The model covariance matrix is the MLE of the sample covariance matrix under the constraint that particular elements of the inverse covariance matrix are zero. The zero elements correspond to variable pairs that are hypothesized to be conditionally independent. The MLE covariance matrix may be easily computed using an iterative algorithm (Wermuth and Scheidt 1977; Speed and Kiiveri 1986). The estimated covariance matrix exactly matches the sample covariance matrix except for those elements corresponding to variable pairs that were specified as conditionally independent.

Model Goodness-of-Fit

The computed MLE of the covariance matrix under the specified conditional independence model is compared to the sample covariance matrix using a measure of fit called ''model deviance'' (Whittaker 1990). Model deviance is defined as:

$$D = N \ln(|\hat{\Sigma}|/|\mathbf{S}|), \tag{7}$$

where N is equal to the number of observations, $\mathbf{S}$ is the observed sample covariance matrix, and $|\hat{\Sigma}|$ is the determinant of the MLE of the covariance matrix (Edwards 1995). Model deviance has an asymptotic χ^2-distribution with degrees of freedom equal to the number of excluded edges (Whittaker 1990). Significant values of D suggest an unsatisfactory fit of the model. Smaller values of D indicate a better fit between the proposed model of conditional independence and the observed data (Krzanowksi and Marriott 1995). The model deviance is a test of the specified model against the saturated model, the model including all possible interactions (edges). This saturated model has a deviance of zero, because it exactly reconstructs the observed covariance matrix. The null hypothesis is, therefore, that all conditional interactions are significantly nonzero. In the context of assessing integration, using the saturated model as the null hypothesis suggests that a relatively large amount of nonzero conditional association among traits is the default expectation (Chernoff and Magwene 1999). The opposite of the saturated model is one in which none of the interactions are significant. A model of complete independence among variates has maximal deviance (Whittaker 1990). Generally speaking, the goal of graphical modeling is to find the model(s) with the fewest interactions that is in accordance with the observed covariance matrix.

The deviance difference for comparing two models, M_0 and M_1, where M_0 is a subset of M_1 is

$$d = N \ln(|\hat{\Sigma}_0|/|\hat{\Sigma}_1|) \tag{8}$$

(Edwards 1995), where $\hat{\Sigma}_0$ is the MLE for the model M_0 and

TABLE 6. Maximum-likelihood-estimate correlation matrix (lower triangle) for the fowl dataset based on the model in Figure 4a. The elements in bold type are those values that are approximated in the construction of the maximum-likelihood estimate. The observed correlations are given in the upper triangle of the matrix. Variables are defined in Table 1.

	L	B	H	U	F	T
L	*	0.583	0.621	0.603	0.569	0.602
B	0.583	*	0.584	0.526	0.515	0.548
H	**0.340**	0.584	*	0.937	0.877	0.874
U	**0.307**	0.526	0.937	*	0.878	0.894
F	**0.293**	**0.503**	0.877	**0.872**	*	0.926
T	**0.290**	**0.497**	**0.874**	0.894	0.926	*

$\hat{\Sigma}_1$ is the same for model M_1. The deviance difference also has an asymptotic χ^2-distribution, with degrees of freedom equal to the difference in the number of edges (Edwards 1995).

The Fowl Data Reconsidered

I return now to a consideration of the fowl skeletal dataset analyzed above. The total deviance in the dataset (the deviance against a model of complete independence) is 1917.47 with 15 degrees of freedom. The model deviance of the graph of Figure 4a is 69.08 with eight degrees of freedom. The model accounts for approximately 96% (1 − 69.08/1917.47) of the total variability in the sample, while requiring many fewer interaction parameters than does the saturated model.

Despite the relatively large proportion of variation summarized by the model, a deviance of 69.08 with eight degrees of freedom indicates an unsatisfactory fit. Edge exclusion deviances alone are, in general, not sufficient to specify a final model with adequate fit (Whittaker 1990). An examination of the MLE correlation matrix (Table 6) suggests why this is the case. All of the estimated correlations (bold, lower triangle) are fairly close to the observed correlations (upper triangle) except for the last four values of the first column. The model, as currently formulated, does not constrain the MLE well enough to adequately recreate the correlations of head length with the limb measures. Referring to the previous matrix of edge exclusion deviances suggests that the addition of either {L, H} or {L, T} may improve the fit of the model. Adding {L, H} seems the more logical given the previous considerations. The new model (Fig. 4d) has a deviance of 13.99. The deviance difference between the new model and the previous one is 55.09, which is significant. A deviance of 13.99 on seven degrees of freedom fits sufficiently well relative to the saturated model.

The new model differs slightly from the previous one, but biological interpretation of the model is similar. There is still an indication of head, wing, and leg compartments, with somewhat weaker proximal and distal limb compartments. The interactions between head and wing variables remain relatively weak despite the addition of an edge in the conditional independence graph.

In fact, the addition of any additional edge involving head length, except for the edge {L, F}, results in a satisfactory fit. This suggests that the partial correlation coefficients of head length with limb measures are close to, but not quite zero. Based on additional information about edge deviances and edge strengths, one might argue that head length is, for all practical purposes, conditionally independent of the limb measures, and accept the model of Figure 4b. Alternately one might argue that the weak edges included in Figure 4d reflect a significant, if small, interaction between head and forelimb measures. Further empirical evidence (morphometric, genetic, developmental) about the traits in question is needed to argue strongly for one model over the other. Statistical methods are unlikely to suggest single answers when the systems under consideration are biologically complex enough to be interesting. Empirical data, theory, and good sense must be one's guides. Below I present another example in which I test specific a priori hypotheses using a graphical modeling approach.

A Second Example: Berg's Correlation Pleiades

Berg (1960) examined correlation matrices for 19 herbaceous plant species and concluded that plants with specialized insect pollinators exhibited patterns of association in which reproductive traits and vegetative traits formed distinct subsets (correlation pleiades). Traits within either subset showed generally high levels of correlation with other traits in the same subset, although correlations between the subsets were weak. The distinct patterns of correlation which Berg observed for taxa with specialized pollinators contrasted sharply with the patterns for plants that were self-pollinated, wind-pollinated, or pollinated by unspecialized insects. Plants lacking specific pollinators exhibited a pattern of trait interaction in which all traits were closely intercorrelated and thus failed to show evidence of correlation pleiades.

In her analysis, Berg (1960) included a number of representative correlation matrices for species in which vegetative and reproductive modules were observed and others in which correlation pleiades were absent. I reexamined Berg's data using graphical models to test the support for these two categories. As representative of a species exhibiting correlation pleiades, I reanalyzed the published data for *Chamaenerium angustifolium* (fireweed; bee-pollinated); *Hordeum vulgare* (barley; self-pollinated) was taken to represent a lack of correlation pleiades. (Note: The published correlation matrix for *Chamaenerium* is singular, probably due to roundoff. To work with an invertible matrix, I estimated a ''bent'' covariance matrix from the original matrix. This bending was achieved by setting the single negative eigenvalue of the original matrix to a small positive value and recalculating a bent covariance matrix as $\mathbf{S}' = \mathbf{E}\Lambda'\mathbf{E}^{\mathrm{T}}$, where Λ' represents the modified diagonal matrix of eigenvalues, and $\mathbf{E}$ represents the original matrices of eigenvectors. This procedure seems to introduce relatively little bias to the elements of the estimated correlation matrix. The original and bent correlation matrices are quite similar, with a maximum elementwise difference of 0.037 between the two matrices.)

Figure 6a shows the graphical model estimated for *Hordeum.* This model has very good fit—a model deviance of 13.2 with 15 degrees of freedom—and summarizes approximately 98% of the sample variability. As Berg suggested, there is no strong evidence for distinct reproductive and vegetative pleiades for this species. However, neither is there

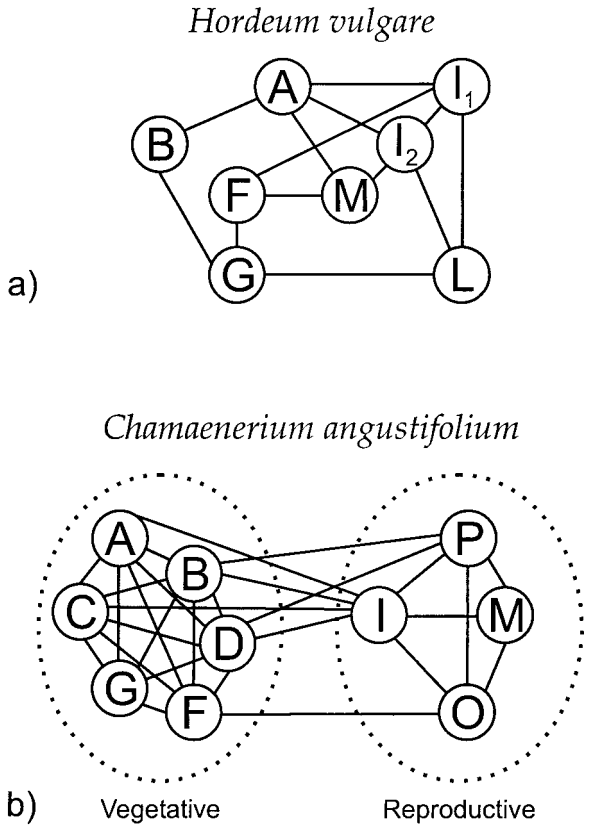

FIG. 6. Independence graphs for plant parts based on analysis of data from Berg (1960). (a) Independence graph for *Hordeum vulgare*; (b) independence graph for *Chamaenerium angustifolium.* A, stem height; B, leaf length; C, leaf width; D, leaf number; F, inflorescence length; G, number of florets (or spikelets); I, flower diameter (I_1, length of lemma; I_2, length of palea); L, grain length; M, sepal length (glume length); O, stamen length; P, pistil length. See text for further discussion.

good evidence for a single large subset in which all traits interact with each other. The model of Figure 6a suggests that less than half (13 of 28) of the pairwise interactions are significant.

An independence model for *Chamaenerium* is shown in Figure 6b. The MLE of the covariance matrix for the given independence model has a deviance of 22.5 with 17 degrees of freedom and captures approximately 95% of the total variability in the data while eliminating approximately 38% (17 of 45) of the pairwise interactions among traits. This model provides support for Berg's interpretation of distinct vegetative and reproductive pleiades, with relatively strong interactions within modules and fewer interactions between modules.

INTEGRATION AND MODULARITY

With a set of statistical and analytical tools for exploring interactions in hand, I now turn to the fundamental question of how to define phenotypic modules: integrated subsets of traits.

The general notion expressed in the literature is that morphological integration is indicated by patterns of association (usually correlations or covariances) that are stronger among some subsets of traits than between others (Olson and Miller 1958; Cheverud 1982). The level of association taken as indicative of integration may be arbitrarily defined (Olson and Miller 1958) or, more typically, is left unstated (e.g., Cheverud 1982, 1995; Zelditch 1987, 1988). In some sense, models which test a priori hypotheses of integration (Cheverud 1982, 1995; Zelditch 1987, 1988), avoid the problem of rigorously defining integrated subsets. If the model adequately fits the data, then the specified subsets are deemed to be integrated. In part, the lack of an explicit definition of integration in previous discussions reflects the biological reality at hand, namely that integration is a matter of degree. Acknowledging inherent biological fuzziness does not, however, preclude us from establishing explicit criteria that are useful for modeling biological systems. Models are, by definition, simplifications of complex situations. Generally one is willing to accept the simplifying dichotomies, boundaries, and distinctions of a model if they lead to new insights and predictions about the system of interest. I show below that a definition of integration based on conditional independence relationships leads to such insights (see the section on multivariate selection).

Closely allied to the study of morphological integration is a literature concerned with the modular nature of genotypes and phenotypes (Frazzetta 1975; Riedl 1978; Bonner 1988; Atchley and Hall 1991; Wagner 1995, 1996; Raff 1996; Wagner and Altenberg 1996; Bolker 2000). The concept of modularity encapsulates the notion that subsets of traits act in a semi-autonomous fashion over development or evolution. Most often, hypotheses of modularity are suggested by phenotypic patterns, but the evolution of modularity must ultimately rest in the evolution of modular genetic systems, reflected in patterns of pleiotropic interaction (Cheverud 1984; Bonner 1988; Wagner and Altenberg 1996). Modularity is attested to by the dissociability of developmental processes (Needham 1933; Gould 1977; Raff 1996) and coordinated patterns of trait evolution across wide taxonomic groups (Wagner 1995).

Wagner (1996, p. 38) provided three criteria for recognizing modular phenotypic units: A phenotypic module ''is a complex of characters that (1) collectively serve a primary functional role; (2) are tightly integrated by strong pleiotropic effects of genetic variation; and (3) are relatively independent from other such units.'' Wagner and Altenberg (1996) further elaborated on these criteria. In particular, they noted that because of the complex interactions among phenotypic traits, a useful analytical method for studying modularity would have to allow for overlapping, hierarchical, and gradual patterns of interaction among modules. Wagner's (1996) first criterion can be modified somewhat to be more inclusive; a module corresponds to a particular biological criterion (Chernoff and Magwene 1999). Modules need not do something; they may just as legitimately be the result of some process, for example, a byproduct of the developmental rules for building particular morphological structures. Bolker (2000) provides an extensive discussion of alternate definitions of modularity, primarily from a developmental perspective.

Integrated morphological subsets and modular units of the phenotype are synonymous. In the following discussion, the term ''module'' will serve as a convenient shorthand for

''integrated subset.'' The definition of modularity that I propose below is explicitly statistical in nature and is based on population-level patterns of covariation.

What Constitutes a Module?

I now present an analytical definition for modular trait complexes. This definition is based on the concept of conditional independence, described in previous sections, and the analytical recognition of patterns of conditional independence among sets of traits. The definition is applicable to both continuous and discrete traits, although I focus on the continuous case. Both strong and weak definitions of modularity are given.

Strong definition.—A module is a maximal subset of traits for which *all* pairs of traits within the subset are mutually informative, conditional on all other traits under consideration.

The term ''informative'' is used here in the sense of information theory. Classical information theory is concerned with measuring the information in one random variable about the value of another (Pierce 1980; Whittaker 1990). Information proper is related to the problem of prediction, of ''average predictability'' (Whittaker 1990).

The insistence on conditional informativeness results from the observation that, for most biological datasets, there are generally high overall phenotypic correlations among traits. Much of this correlation may be due to joint associations with other variables, either observed or unobserved. By focusing on conditional association, the definition emphasizes interactions that provide the greatest amount of information about patterns of variability, while ignoring those interactions that contribute little or nothing to further knowledge about the systems under study.

The above definition of modularity has a very specific interpretation in the context of an independence graph based on covariances or correlations among phenotypic traits. The edges of an independence graph based on phenotypic traits indicate which pairs of traits are mutually informative. In such a graph, modules correspond to the cliques of the graph. A clique is a graph subset that is complete (i.e., all vertices in the subset are adjacent) and maximal (i.e., adding any other vertex to a clique implies that not all vertices are adjacent; Lauritzen 1996).

Cliques may overlap, mirroring the complexity inherent in biological systems. The cliques of the graph in Figure 4a are {L, B}, {B, H, U}, {F, T}, {H, F} and {U, T}, corresponding to head, head + wing, leg, and proximal and distal modules, respectively. The modules of integration, as defined above, are shown in Figure 4e. The head, wing, and leg modules are consistent with Olson and Miller's (1951, 1958) notion that functionally related traits should exhibit patterns of integration, whereas the weaker interactions between head and wing traits probably reflect topological proximity rather than any explicit functional demands. The proximal and distal modules, in contrast, are best interpretted as reflecting developmental interactions relevant to the patterning of the limbs.

Although cliques can overlap, they cannot form nested (hierarchical) sets. However, patterns of interaction within cliques may show hierarchical structure. Particular subsets of traits within modules may show stronger patterns of interaction as indicated by varying degrees of edge strengths (Fig. 4b). A weaker definition of modularity that also allows for hierarchical patterns among modules is as follows.

Weak definition.—A module is a subset of traits for which *most* pairs of traits within the subset are mutually informative, conditional on all other traits under consideration.

This weak definition simply relaxes the criterion of completeness by allowing for a small number of trait pairs within a module to be noninformative about each other. In the case of the fowl dataset, one might recognize a general limb module in which wing, leg, proximal, and distal subsets are nested. What constitutes a small number is obviously subjective.

Size as a Factor

It is commonly assumed that the strongest source of association among morphometric variables is due to the nature of growth processes. As a general rule, organismal structures get larger as ontogeny proceeds. Joint correlations with growth may largely obscure the specific patterns of integration sought in correlation or covariance matrices. In the context of exploratory path analyses, a common maneuver is to first fit a single factor model to the data, where the factor is assumed to represent general size. Large residual correlations remaining after fitting a single factor model are taken to indicate strong associations not related to size (Wright 1932; Bookstein et al. 1985).

In the context of the graphical modeling approach described above, there seems to be no explicit allowance made to incorporate such size factors. In fact, in the case of continuous morphometric variables, size is implicitly estimated. The elements of the scaled inverse covariance matrix correspond to the partial correlations of the respective variables, conditional on all the rest. When the measured traits are continuous morphometric variables, ''all the rest'' can be construed to be an estimate of general size based on all the variables except the two under consideration. Whittaker (1990, p. 260) presents an example that provides a similar line of reasoning.

Integration and Multivariate Selection

The definition of integration presented above provides a useful set of statistical criteria for defining integrated subsets of traits. Conditional independence relationships as assessed from population data also have implications in the context of standard models of multivariate selection.

Lande and Arnold (1983) showed how a retrospective approach could be used to infer patterns of directional and stabilizing selection on phenotypically correlated characters. The vector $\boldsymbol{\beta}$, the vector of directional selection gradients, characterizes directional selection, and is equivalent to the product $\mathbf{P}^{-1}\mathbf{s}$ where $\mathbf{P}$ is a matrix of phenotypic correlations and $\mathbf{s}$ is a vector of selection differentials. Conditional independence among a pair of traits represents a special case in which selection on one trait does not affect relative fitness with respect to the other trait (this does not imply, however, that selection on one will not cause a change in the mean of the other).

For example, assume that there are three traits, X, Y, and Z, that are relevant for determining fitness. Furthermore, assume X and Z are independent conditional on Y, and an arbitrary directional selection vector. The conditional covariance matrix and vector of selection differentials will have the form:

$$\mathbf{P}^{-1} = \begin{bmatrix} d_{11} & d_{12} & 0 \\ d_{21} & d_{22} & d_{23} \\ 0 & d_{32} & d_{33} \end{bmatrix} \quad \text{and} \tag{9a}$$

$$\mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix}. \tag{9b}$$

Therefore, $\boldsymbol{\beta} = \mathbf{P}^{-1}\mathbf{s} = (s_1d_{11} + s_2d_{22}, s_1d_{21} + s_2d_{22} + s_3d_{33}, s_2d_{32} + s_3d_{33})^{\mathrm{T}}$. This shows that when X and Z are conditionally independent, selection on X has no effect on relative fitness with respect to Z and vice versa.

These results lend support to the definition of integration given above. Traits which are unintegrated are essentially invisible to each other with respect to their relative fitness effects. The implications of conditional independence in the context of multivariate selection will be discussed further in a subsequent report (P. M. Magwene and S. H. Rice, unpubl. ms.)

Discussion

The method presented here for exploring patterns of trait interaction and defining modules of integration is in the same spirit as a number of other approaches to studying trait interaction, particularly exploratory path analysis (Wright 1932; Bookstein et al. 1985). Why advocate for the adoption of a new approach?

It is important to recognize that studying morphological integration is not an either-or proposition. The tools and techniques of various approaches may compliment each other at different stages of investigation. For example, analyses such as those advocated here may suggest the existence of latent variables, which can be tested in the context of confirmatory factor analyses. The approaches that have been advocated by various authors (e.g., Van Valen 1965; Zelditch 1987, 1988; Cheverud et al. 1989) are, for the most part, complimentary.

An approach based on graphical modeling stands out for a number of reasons. Most compelling is the fact that conditional covariance among trait is related to directional selection gradients. By the definition given above, a module is a set of traits that all have relative fitness effects on each other. One searches for such sets by discovering those pairs or sets of traits that are conditionally independent, and thus have no relative fitness interaction.

Another advantage is that the technique recommended here is relatively simple, yet provides a variety of statistics (edge exclusion deviance, edge strength, model deviance) for exploring the fit of models as a whole, as well as specific interactions among traits. The simplicity of the approach is self-evident. Useful information about trait interactions can be gained by simply inverting the covariance or correlation matrix. The robustness and strength of single edges in an independence graph can be summarized using edge exclusion deviances and edge strengths, which are also very simply calculated. The most complicated step, algorithmically, relates to model testing, and involves the calculation of MLEs of the covariance matrix based on specified patterns of conditional independence. The algorithm presented by Wermuth and Scheidt (1977) is relatively easy to implement. Other equally simple approaches to the study of morphological integration such as those based on cluster analyses (Van Valen 1965; Cheverud 1982) provide no well-defined goodness-of-fit measures. The matrix correlation approach (Cheverud et al. 1989) can provide a goodness-of-fit measure for a model as a whole (Dow and Cheverud 1985), but there is no complimentary statistic for assessing the robustness or strength of single interactions.

Edge strengths may serve as testable hypotheses to be explored in selection experiments or developmental manipulations. As a measure of the strength of interaction among traits, edge strengths can be viewed as hypotheses about how easy or difficult it is to disrupt modules of integration. Edges with large edge strengths should require strong selection or fairly drastic alterations to development to disrupt interactions among traits; weak edges should be relatively easier to dissolve. The results of experiments designed to alter or disturb integrative patterns may be checked against expectation using some of the statistical tools outlined above. Perhaps most importantly, carefully designed experiments can lend support for causal hypotheses of integration or lack thereof, such as common responses to local morphogens, functional interaction, or competition for developmental resources (e.g., Nijhout and Emlen 1998).

In addition to their simplicity and the powerful tools for exploring interactions that they provide, the techniques of graphical modeling and the associated depiction of hypotheses via independence graphs are complimentary to the definition of modularity presented above. As Wagner and Altenberg (1996) have argued, a clear and quantitative definition of modularity is essential for exploring the modular nature of genotypes and phenotypes.

Conclusion and Summary

The quantitative study of phenotypic integration and the modular nature of phenotypes is largely concerned with describing patterns of trait interaction and delimiting those subsets of traits that exhibit strong intercorrelations. The notion of conditional independence and the tools of graphical modeling provide a logical basis for exploring patterns of interaction. Independence graphs and measures of information make possible an explicit definition of modularity that is clear and concise and allows for overlapping and hierarchical modules. Combining these tools facilitates both descriptive and experimental studies of phenotypic integration by enabling both heuristic exploration as well as hypothesis testing. Studies that combine descriptive and experimental approaches for studying trait interactions will provide valuable information for furthering our understanding of the modular nature of organismal phenotypes and their underlying genetic representations.

Acknowledgments

This manuscript benefitted from insightful comments and critiques provided by M. LaBarbera, B. Chernoff, S. Rice, G. Wagner, N. Johnson, J. Kim, L. Van Valen, and M. Zelditch. This work was partially supported by a National Science Foundation Dissertation Improvement Grant (M. LaBarbera, PI) and a Ford Foundation Dissertation Fellowship for Minorities.

Literature Cited

Atchley, W. R., and B. K. Hall. 1991. A model for development and evolution of complex morphological structures. Biol. Rev. 66:101–157.

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

Bolker, J. A. 2000. Modularity in development and why it matters to Evo-Devo. Am. Zool. 40:770–776.

Bonner, J. T. 1988. The evolution of complexity. Princeton Univ. Press, Princeton, NJ.

Bookstein, F. L., B. Chernoff, R. Elder, J. Humphries, G. Smith, and R. Strauss. 1985. Morphometrics in evolutionary biology. Academy of Natural Sciences, Philadelphia, PA.

Cane, W. P. 1993. The ontogeny of postcranial integration in the common tern, *Sterna hirundo.* Evolution 47:1138–1151.

Chernoff, B., and P. M. Magwene. 1999. Morphological integration: forty years later. Pp. 316–360 *in* E. C. Olson and R. L. Miller, eds. Morphological integration. Univ. of Chicago Press, Chicago, IL.

Cheverud, J. M. 1982. Phenotypic, genetic, and environmental morphological integration in the cranium. Evolution 36:499–516.

———. 1984. Quantitative genetics and developmental constraints on evolution by selection. J. Theor. Biol. 110:155–172.

———. 1995. Morphological integration in the saddle-back tamarin (*Saguinus fuscicollis*) cranium. Am. Nat. 145:63–89.

Cheverud, J. M., J. J. Rutledge, and W. R. Atchley. 1983. Quantitative genetics of development: genetic correlations among age-specific traits values and the evolution of ontogeny. Evolution 37:895–905.

Cheverud, J. M., G. P. Wagner, and M. M. Dow. 1989. Methods for the comparative analysis of variation patterns. Syst. Zool. 38:201–213.

Cheverud, J. M., E. J. Routman, and D. J. Irschick. 1997. Pleiotropic effects of individual gene loci on mandibular morphology. Evolution 51:2006–2016.

Cox, D. R., and N. Wermuth. 1996. Multivariate dependencies: models, analysis, and interpretation. Chapman and Hall, New York.

Darroch, J. N., S. L. Lauritzen, and T. P. Speed. 1980. Markov fields and log linear interaction models for contingency tables. Ann. Stat. 8:522–539.

Dawid, A. P. 1979. Conditional independence in statistical theory. J. R. Statist. Soc. B 41:1–31.

Dempster, A. P. 1972. Covariance selection. Biometrics 28: 157–175.

Dow, M. M., and J. M. Cheverud. 1985. Comparison of distance matrices in studies of population structure and genetic microdifferentiation: quadratic assignment. Am. J. Phys. Anthropol. 68:367–374.

Edwards, D. 1995. Introduction to graphical modelling. Springer-Verlag, New York.

Frazzetta, T. H. 1975. Complex adaptations in evolving populations. Sinauer Associates, Sunderland, MA.

Gould, S. J. 1977. Ontogeny and phylogeny. Harvard Univ. Press, Cambridge, MA.

Krzanowksi, W. J., and F. H. C. Marriott. 1995. Multivariate analysis. Pt. 2. Classification, covariance structures and repeated measurements. Arnold, London.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain-body size allometry. Evolution 33: 402–416.

Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution 37:1210–1226.

Lauritzen, S. L. 1996. Graphical models. Oxford Univ. Press, New York.

Marcus, L. F. 1990. Traditional morphometrics. Pp. 77–122 *in* F. J. Rohlf and F. L. Bookstein, eds. Proceedings of the Michigan Morphometrics Workshop. University of Michigan Museum of Zoology, Ann Arbor, MI.

Needham, J. 1933. On the dissociability of the fundamental process in ontogenesis. Biol. Rev. 8:180–233.

Nemeschkal, H. L. 1999. Morphometric correlation patterns of adult birds (Fringillidae: Passeriformes and Columbiformes) mirror the expression of developmental control genes. Evolution 53: 899–918.

Nijhout, H. F., and D. J. Emlen. 1998. Competition among body parts in the development and evolution of insect morphology. Proc. Natl. Acad. Sci. 95:3685–3689.

Olson, E. C., and R. L. Miller. 1951. A mathematical model applied to a study of the evolution of species. Evolution 5:325–338.

———. 1958. Morphological integration. Univ. of Chicago Press, Chicago, IL.

Pierce, J. R. 1980. An introduction to information theory: symbols, signals and noise. Dover Publications, New York.

Raff, R. A. 1996. The shape of life. Univ. Chicago Press, Chicago, IL.

Rice, S. H. 2000. The evolution of developmental interactions: epistasis, canalization, and integration. Pp. 82-98 *in* J. B. Wolf, E. D. Brodie, and M. J. Wade, eds. Epistasis and the evolutionary process. Oxford Univ. Press, New York.

Riedl, R. 1978. Order in living organisms. John Wiley and Sons, New York.

Sokal, R. R. 1962. Variation and covariation of characters of alate *Pemphigus populi-transversus* in eastern North America. Evolution 16:227–245.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry. W. H. Freeman, New York.

Speed, T. P., and H. T. Kiiveri. 1986. Gaussian Markov distributions over finite graphs. Ann. Statist. 14:138–150.

Stirzaker, D. 1994. Elementary probability. Cambridge Univ. Press, New York.

Van Valen, L. 1965. The study of morphological integration. Evolution 19:347–349.

Wagner, G. P. 1995. The biological role of homologoues: a building block hypothesis. Neves Jahrb. Geol. Palaeontol. Abh. 19: 279–288.

———. 1996. Homologues, natural kinds and the evolution of modularity. Am. Zool. 36:36–43.

Wagner, G. P., and L. Altenberg. 1996. Complex adaptations and the evolution of evolvability. Evolution 50:967–976.

Wermuth, N., and E. Scheidt. 1977. Fitting a covariance selection model to a matrix: algorithm AS 105. Appl. Stat. 26:88–72.

Whittaker, J. 1990. Graphical models in applied mathematical multivariate statistics. John Wiley and Sons, New York.

Wright, S. 1918. On the nature of size factors. Genetics 3:367–374.

———. 1921. Correlation and causation. J. Agric. Res. 20:557–585.

———. 1932. General, group and special size factors. Genetics 15: 603–619.

———. 1934. The method of path coefficients. Ann. Math. Stat. 5:161–215.

———. 1954. The interpretation of multivariate systems. Pp. 11–33 *in* O. Kempthorne, T. A. Bancroft, J. W. Gowen, and J. L. Lush, eds. Statistics and mathematics in biology. Hafner Publishing, New York.

———. 1968. Evolution and the genetics of populations. Univ. of Chicago Press, Chicago, IL.

Zelditch, M. L. 1987. Evaluating models of developmental integration in the laboratory rat using confirmatory factor analysis. Syst. Zool. 36:368–380.

———. 1988. Ontogenetic variation in patterns of phenotypic integration in the laboratory rat. Evolution 42:28–41.

Corresponding Editor: M. Zelditch