Syst. Biol. 0(0):1–13, 2021

DOI:10.1093/sysbio/syab088

On Information Rank Deficiency in Phenotypic Covariance Matrices

F. Robin O'Keefe[1,*], Julie A. Meachen[2], and P. David Polly[3]

[1]*Department of Biological Sciences, Marshall University, One John Marshall Drive, Huntington, WV 25701, USA;* [2]*Department of Anatomy, Des Moines University, 3200 Grand Avenue, Des Moines, IA 50312, USA; and* [3]*Department of Earth and Atmospheric Sciences, Indiana University, 1001 E 10th St, Bloomington, IN 47408, USA*

**Correspondence to be sent to: Biological Sciences, Marshall University, College of Science 265, One John Marshall Drive, Hunington, WV 25755, USA; E-mail: okeefef@marshall.edu.*

Received 16 October 2020; accepted 27 October 2021
Associate Editor: Lauren Esposito

Abstract.—This article investigates a form of rank deficiency in phenotypic covariance matrices derived from geometric morphometric data, and its impact on measures of phenotypic integration. We first define a type of rank deficiency based on information theory then demonstrate that this deficiency impairs the performance of phenotypic integration metrics in a model system. Lastly, we propose methods to treat for this information rank deficiency. Our first goal is to establish how the rank of a typical geometric morphometric covariance matrix relates to the information entropy of its eigenvalue spectrum. This requires clear definitions of matrix rank, of which we define three: the full matrix rank (equal to the number of input variables), the mathematical rank (the number of nonzero eigenvalues), and the information rank or "effective rank" (equal to the number of nonredundant eigenvalues). We demonstrate that effective rank deficiency arises from a combination of methodological factors—Generalized Procrustes analysis, use of the correlation matrix, and insufficient sample size—as well as phenotypic covariance. Secondly, we use dire wolf jaws to document how differences in effective rank deficiency bias two metrics used to measure phenotypic integration. The eigenvalue variance characterizes the integration change incorrectly, and the standardized generalized variance lacks the sensitivity needed to detect subtle changes in integration. Both metrics are impacted by the inclusion of many small, but nonzero, eigenvalues arising from a lack of information in the covariance matrix, a problem that usually becomes more pronounced as the number of landmarks increases. We propose a new metric for phenotypic integration that combines the standardized generalized variance with information entropy. This metric is equivalent to the standardized generalized variance but calculated only from those eigenvalues that carry nonredundant information. It is the standardized generalized variance scaled to the effective rank of the eigenvalue spectrum. We demonstrate that this metric successfully detects the shift of integration in our dire wolf sample. Our third goal is to generalize the new metric to compare data sets with different sample sizes and numbers of variables. We develop a standardization for matrix information based on data permutation then demonstrate that *Smilodon* jaws are more integrated than dire wolf jaws. Finally, we describe how our information entropy-based measure allows phenotypic integration to be compared in dense semilandmark data sets without bias, allowing characterization of the information content of any given shape, a quantity we term "latent dispersion". [*Canis dirus*; Dire wolf; effective dispersion; effective rank; geometric morphometrics; information entropy; latent dispersion; modularity and integration; phenotypic integration; relative dispersion.]

The constituent parts of an organism form an integrated whole, and it is this integrated whole that develops ontogenetically and is subject to evolutionary change in populations (Olson and Miller 1958). This notion of integration in biological systems can be measured as the degree of correlation among the parts of an organism and can refer to phenotype, genotype, and other factors (Cheverud 1982). In this classic paper, Cheverud demonstrated that the functional units of the macaque cranium were tightly correlated and are therefore not free to develop, or evolve, independently. This correlation, termed "phenotypic integration" when referring to shape, is critical in evolving populations because traits are not free to respond to selection without impacting dependent traits, and the directionality of selection response is constrained by these dependencies (Grabowski and Porto 2017). Yet trait dependency can also remove constraint by giving a population access to novel areas of adaptive space (Goswami et al. 2014, Fig. 5). Consequently, the integration of phenotypic traits is central to the evolvability of biological systems, and the topic has received much scrutiny (see Pavlicev et al. 2009a, for a review).

The use of landmark data to analyze biological shape has become ubiquitous since the modern codification of geometric morphometrics by Bookstein (1997); reviewed in Zelditch et al. (2012). Geometric morphometric data have long been used to study phenotypic integration, or the strength and patterns of covariation among the morphological features comprising a shape (e.g., Klingenberg 2008, Klingenberg 2013; Klingenberg and Zaklan 2000; Zelditch et al. 1992). The strength of phenotypic integration is frequently measured by characterizing covariance patterns found in a covariance matrix of landmarks, most often with reference to the distribution of the matrix's vector of eigenvalues (e.g., Wagner 1984; Pavlicev et al. 2009a,b). This eigenvalue spectrum contains information about phenotypic integration because, as integration increases, variance is more concentrated on the first few eigenvectors. An eigenvalue spectrum with no integration would have a flat scree plot while increasing integration leads to larger initial values and smaller later values in the scree plot. A traditional integration metric like the eigenvalue standard deviation measures this property of the eigenvalue spectrum. It is known that the covariance matrix can be rank deficient due to superimposition with Generalized Procrustes Analysis (GPA), and when there are fewer specimens than variables (Dryden and Mardia 1998; Rohlf 1990).

However, there is another type of rank deficiency that has not been fully explored: this is the amount of information redundancy in the eigenvalue spectrum. In this article, we demonstrate that this information rank deficiency is high, and that it adversely impacts integration measures, in a model system of dire wolves. New measures are then proposed that account for information redundancy. The aim is to measure integration with respect to the information content of a shape, not with respect to the (arbitrary) number of input variables used to measure it.

Phenotypic Integration Measures

"The 'generalized variance', $|\Sigma|$, the determinant of the variance-co-variance matrix, is related to the area (or hypervolume) of the equi-probability ellipses (ellipsoids) of the distribution."—VanValen (1974, p. 235).

Properties of the distribution of the eigenvalue vector, such as its mean or standard deviation, form the basis of several metrics intended to quantify the strength of phenotypic integration. As stated by VanValen, the shape space containing the objects of interest can be conceptualized as a hyperellipse in the original variable space. Principal components analysis moves this hyperellipse to the origin, rotates it, and defines its principal axes. VanValen realized that two aspects of this ellipse are pertinent to measuring its degree of integration: i) its dispersion in dimensionality and ii) its dispersion in variance on each axis, which he called "tightness." While variance dispersion has received a great deal of attention in subsequent literature, dimensionality dispersion has not (reviewed in Najarzadeh 2019).

Cheverud et al. (1983) introduced the first modern measure of phenotypic integration, defining it as

$$I = 1 - \left(\prod_{i=1}^{v} \lambda_i\right)^{1/v}$$

or the geometric mean of the eigenvalues of the correlation matrix of variables v (not the covariance matrix) subtracted from unity. This quantity is identical to one minus the vth root of the generalized variance (Wilks 1932), the quantity proposed by VanValen 1974. In a wider statistical context, this quantity is termed the "standardized generalized variance," or SGV (SenGupta 1987). The SGV is the geometric mean of the eigenvalues, and so may be thought of as the mean diameter of the axes in the hyperellipse. In this article, we use 14 landmarks, which yield a total of 28 coordinate variables, and so the full rank of the covariance matrix will be 28. However, four degrees of freedom are lost of the Procrustes analysis, yielding a full rank of 24 (see Discussion below). Therefore, the SGV is here calculated as the 24th root of the first 24 eigenvalues for the matrices in this article:

$$\mathrm{SGV}_{24} = \sqrt[24]{\prod_{i=1}^{24} \lambda_i}$$

Both Cheverud et al. (1983) and SenGupta (1987) state that the SGV is comparable (as a measure of dispersion) among spaces of different dimensionality, and this assumption is widely accepted, although it has never been tested (e.g., Najarzadeh 2019). While both the Cheverud integration and the nearly equivalent SVG have been used as measures of dispersion, a consensus has emerged that the standard deviation of the eigenvalues has more desirable statistical properties. This eigenvalue dispersion is measured as the standard deviation of the eigenvalues of the correlation matrix of GPA landmarks, standardized to the mean eigenvalue (Pavlicev et al. 2009a):

$$\lambda\sigma = \frac{\sqrt{\sum_{i=1}^{v}\left(\frac{\lambda_i}{\mu\lambda} - 1\right)^2}}{\sqrt{v-1}}$$

Several other metrics have been proposed for quantifying the dispersion of phenotypic matrices, summarized and compared by Pavlicev et al. (2009a).

Integration and Modularity

"However, if the structure of a dataset is strongly modular, with several different groups of strongly co-varying traits, then variance will be distributed more evenly across a number of principal components, and eigenvalue variance will be relatively low. Thus, the dispersion of eigenvalues provides a simple measure for comparison of the relative integration or modularity of the structures described in a matrix."—Goswami and Polly (2010, p. 226).

In the 21st century, the study of covariance matrices derived from Procrustes superimposition of landmark data has grown to include the field of modularity and integration (Klingenberg 2013; Goswami et al. 2014). This conception of integration is similar to Van Valen's and Cheverud's in that it studies covariance structure but differs in attempting to identify subsets of variables that covary relative to others. Rather than relying on a single overall measure of dispersion, modularity studies usually test specific models of landmark covariation against a null model in an attempt to characterize subsets with high intraset covariation and low interset covariation (Goswami and Polly 2010). Because these sets are primary features of the covariance structure, they should be carried on the first several principal components, and the entire eigenvalue distribution is not directly relevant. This approach to modularity has yielded powerful hypotheses of evolutionary plasticity and constraint linking development to phenotype, and to evolutionary changes in modularity over deep time (e.g., Goswami et al. 2015). Modularity models were first assessed using the RV coefficient introduced by Klingenberg (2008). This coefficient has been superseded by the similar covariance ratio statistic (CR, Adams 2016), which is more robust to differences in sample size and data dimensionality.

In the jargon of the modularity and integration, phenotypic integration is termed "overall modularity and integration," or "whole shape integration" (Goswami and Polly 2010). Determining this quantity is generally not an objective in modularity studies, but when it is measured, the standard eigenvalue dispersion of phenotypic integration is used (Goswami and Polly 2010; Equations 7 and 8). However, phenotypic integration metrics should agree with measures of modularity: a shape should be more integrated (possess fewer modules) if it has greater eigenvalue dispersion, and should be less integrated (possess more modules) if it possesses lesser dispersion, as stated in the quote above. In a population where modularity is evolving, an increase in modularity should lead to a decrease in phenotypic integration and hence eigenvalue dispersion, and thus a decrease in dispersion metrics like SVG and eigenvalue standard deviation.

Rank Deficiency in Phenotypic Covariance Matrices

It [PCA] is used to obtain a more economic description of the N- dimensional dispersion of the original data by a smaller number of "principal components," which are formally the eigenvectors of the dispersion matrix... Hence the number of principal components that contribute significantly to the variation of the sample is the actual "dimensionality" of the dispersion. —G. P. Wagner 1984, pp. 92-93) (emphasis added).

The rank of a matrix is the dimensionality of the vector space spanned by its variables. If all variables are completely independent then the matrix is of "full rank," equal to the number of variables, but when one variable depends completely on another, the matrix has "deficient rank." The full rank value of a matrix is the number of columns; if one estimates a covariance matrix **K** directly from phenotypic traits **X** (without GPA),

$$K_{v_i v_j} = \mathrm{cov}(x_i x_j)$$

the mathematical rank—the number of nonzero eigenvalues in K—will be equal to the full rank. Yet due to the high degree of correlation in phenotypic systems (Van Valen 1974; Cheverud 1982; Adams 2016), many of the higher eigenvalues will be very small. This effect is exacerbated in geometric morphometric matrices because each landmark coordinate is treated as an independent variable, when it is not (demonstrated below). Therefore, geometric morphometric covariance matrices derived from highly integrated systems will have long tails of small, nonzero eigenvalues. This tail should be trivial, and some part of it should probably be ignored, but this phenomenon and its impact have not been explored. In this article, we quantify this tail as another source of rank deficiency, one due to information. Information rank deficiency is defined as rank deficiency arising from trivially small, but nonzero, eigenvalues. Determining the value of information rank deficiency relies on the assessing the information entropy of the eigenvalue spectrum and is formally defined below.

Rank deficiency (meant here in the general sense, including information rank deficiency) arises from at least three sources in geometric morphometric matrices: the GPA, lack of matrix information due to inadequate sample size, and true phenotypic covariance. A fourth possible source is shape distortion introduced by substituting a correlation matrix for the covariance matrix. Of these, only phenotypic covariance is of interest for measuring biological integration, while the others may confound attempts to measure it. The effect of Procrustes analysis on rank deficiency is known. The translation, scaling, and rotation performed during the analysis remove four degrees of freedom from 2D landmarks and seven from 3D landmarks (Zelditch et al. 2012). A covariance matrix derived from 2D GPA coordinate data will therefore possess $v-4$ nonzero eigenvalues, or in other words be mathematically deficient by four ranks from the full rank.

Information rank deficiency resulting from the information content of a landmark covariance matrix is harder to quantify. This lack of information can have two sources: insufficient sample size and landmark oversampling. When the number of specimens n is less than the number of variables v, the maximum rank of the covariance matrix is $n-1$, but matrix rank is affected even when n is greater than v. Grabowski and Porto (2017) studied the effect of sample size on phenotypic integration measures, and concluded that n needed to be more than 10 times v before integration and evolvability metrics are stable to changes in n. Integration studies with smaller sample sizes are therefore information poor, and this should increase information rank deficiency. What we term landmark oversampling refers to the dense landmarking of a shape that varies in a small number of ways. A simple shape with few modes of variation requires few landmarks to characterize fully. Use of more than the necessary number of landmarks will produce redundant covariance among the variables, which increases information rank deficiency. This source of information rank deficiency has not been quantified (but see Bookstein 2015). However, it is tractable, as we outline in the Discussion.

Substituting a correlation matrix for a covariance matrix also affects information rank, by artificially inflating it. In a correlation matrix, each coordinate of each landmark is awarded full rank and given a variance of 1. Normally this standardization allows variables that are measured on different scales to be compared on the same scale. However, Procrustes superimposed landmarks are already standardized to a common shape coordinate system, and the use of a correlation matrix puts each landmark coordinate onto its own independent scale, which warps the variable space by reducing strong directional variation in some coordinates and inflating small directional variance in others (for discussion of the desirability of the covariance matrix for geometric morphometric applications see Goswami and Polly 2010, p. 217).

Information Entropy and Effective Rank

The first to point out the problem of redundancy in the mathematical rank of **K** in a phenotypic integration context was Van Valen (1974), who tailored an information metric to correct for it. The problem of information rank deficiency can be framed in terms of the eigenvalues of a matrix, with the rank of interest being the subset of eigenvalues carrying significant variance, as stated in the quote above. This number is smaller than the set of eigenvalues that carry nonzero variance, and hence smaller than the mathematical rank. The question of the number of "significant" eigenvalues is a general question in data analysis, and there are many techniques for determining this number, summarized by Cangelosi and Goriely (2007) in the context of cDNA microarray data. They include the broken stick model, Cattell's SCREE test, and Bartlett's sphericity test among many others (see also Hine and Blows 2006; Bunea et al. 2011). All these techniques share fundamental shortcomings, such as forcing an integer value onto the rank, and recourse to an *ad hoc* criterion to assess the cutoff for significance.

The problem of redundancy in eigenvalue distributions has been solved by a different strategy in signal processing with "information entropy" (Shannon 1948). Information entropy is a metric used to measure how much a signal (such as an MP3 audio file) can be compressed without loss of information. At least two metrics based on Shannon's information entropy have been proposed to characterize the dimensionality of covariance matrices in a biological context (for a thorough mathematical development see Cangelosi and Goriely 2007). Those authors develop a dimensionality metric for cDNA microarray data they termed the "information dimension." A very similar metric called "effective rank" (and "effective dimensionality") was developed by Roy and Vetterli (2007). The advantages of this approach are that it measures the information content of only the nonredundant nonzero eigenvalues, and it uses a continuous rather than integer scale.

We follow Roy and Vetterli (2007) in our application of Shannon entropy to the information rank problem in geometric morphometrics. Given a covariance matrix **K** of full rank v derived from Procrustes shape coordinates with eigenvalues $\Lambda_{\mathbf{V}}$:

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_v \geq 0$$

following Shannon (1948), the Shannon entropy E_S of **K** is defined as

$$E_s = -1 * \sum_{i=1}^{v} \left(\frac{\lambda_i}{\sum \Lambda_V}\right)\left(ln\frac{\lambda_i}{\sum \Lambda_V}\right).$$

Note that the eigenvalues are standardized to the trace of Λ; this is necessary because the information entropy was originally defined in a probability context, and the terms being evaluated must therefore sum to unity (Shannon 1948). Roy and Vetterli define their effective rank as e raised to the power E_s; the "effective rank" of **K** is here defined in the same way, and is equivalent to the concept of "information rank" discussed above:

$$R_e = e^{Es}.$$

This effective rank, R_e, of **K** is a continuous metric, is based on solid theoretical grounds from information theory, and may be thought of as the significant number of dimensions of the shape space represented by **K**, or equivalently as the number of nonredundant eigenvalues in **K**.

Study Design

In this article, we show that effective rank is a good measure of the nontrivial rank of a geometric morphometric covariance matrix, and we demonstrate its effectiveness on two samples of Pleistocene dire wolf mandibles from Rancho La Brea. We develop our argument in three steps, and introduce new materials and method descriptions at each stage. This organization allows sequential presentation of results as they become necessary for further methodological development.

We first apply the concept of effective rank to investigate the information rank deficiency in landmark data from jaws of two dire wolf populations separated by 5000 years. We demonstrate that the covariance matrices are highly effective rank deficient. Through use of a permutation test, we characterize the magnitude of the effective rank deficiency arising from lack of matrix information. When combined with the rank deficiency expected from the GPA, this allows us to quantify the information rank deficiency due to phenotypic covariance. Lastly, we quantify the rank inflation arising from use of the correlation matrix. ii) We then evaluate the impact of the observed effective rank deficiency on the assessment of integration and modularity changes between the populations. Existing phenotypic integration metrics fail to capture a demonstrable change in integration. We employ jackknifing to produce confidence intervals for hypothesis testing and normalize the rank degeneracy due to matrix information content with a measure we term "effective dispersion." iii) Finally, we generalize the concept of effective dispersion to the case of arbitrary matrix information content, using a metric we call "relative dispersion." This allows comparison of phenotypic integration among data sets of different sample sizes and landmark numbers. We demonstrate its utility by showing that a *Smilodon* data set is much more tightly integrated than either of the dire wolf populations.

Materials and Methods

Two terminal Pleistocene populations of dire wolves (*Canis dirus*) were sampled from two pits (13 and 61/67) at Rancho La Brea. The samples are separated by about 5,000 years (Stock and Harris 1992; Binder and Van Valkenburgh 2010; O'Keefe et al. 2014; Brannick et al. 2015). The material is stored at the Tar Pit Museum

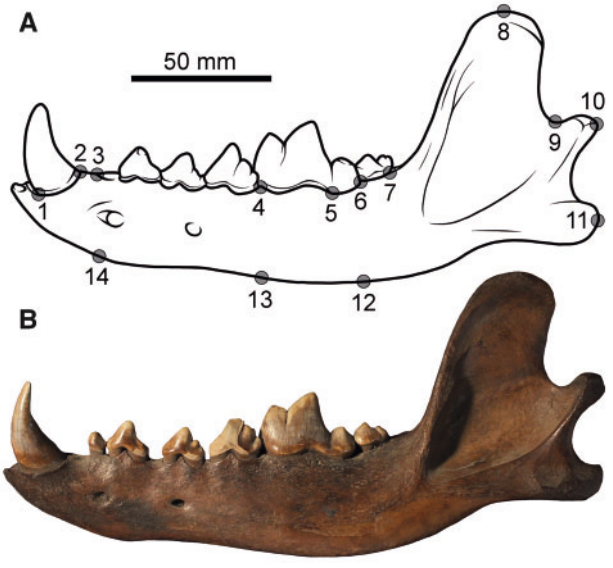

FIGURE 1. Landmarks used in this study, numbered on a schematic dire wolf jaw (a), and (b) a representative *Canis dirus* dentary from Pit 61/67, Rancho La Brea, California.

at Hancock Park, Los Angeles. Pit 13 wolves show elevated tooth breakage and wear, while those from 61/67 do not (Binder et al. 2002), which are thought to indicate differences in nutrient stress at times when prey resources are scarce (Van Valkenburgh 1988, Van Valkenburgh 2009; Meloro 2012). Size and shape differences between the two populations have also been documented and attributed to differences in nutrition (O'Keefe et al. 2014; Brannick et al. 2015; for a full discussion see Supplementary Appendix 2).

Using 119 dire wolf jaws (n= 36 from Pit 13 and n=83 from pit 61/67), we collected 14 landmarks of the 16 used Brannick et al. (2015, Fig. 1) using tpsDig2 (Rohlf 2013). Positions of landmarks were chosen to give a general outline of the mandible and capture information of functional relevance. We omitted landmarks used by Brannick et al. at the angle of the mandible because it was difficult to identify and showed unacceptably large variance, and an interior landmark in the masseteric fossa to make the dataset amenable to outline-based iterative semilandmarking. All specimens were anatomical lefts, laid flat and photographed with a 5 cm scale-bar that allowed us to standardize scale before landmark digitization. We also collected 14 similar landmarks from 81 *Smilodon fatalis* jaws from pits 61/67 and 13 at the Tar Pit Museum and University of California Museum of Paleontology (pooling was necessary to increase sample size).

ANALYSIS AND RESULTS

Part 1: Rank Deficiency in Dire Wolf Mandibles

In this section, we identify the sources of rank deficiency in the covariance matrix derived from the *Canis dirus* data from Pit 61/67. After Procrustes superimposing the landmark data using the **geomorph** package in R (Adams et al. 2020), we calculated the eigenvalues of the correlation and covariance matrices of the superimposed landmarks and correlation and covariance matrices for data from which interlandmark covariance was removed by random permutation of the columns in the original data matrix. Figure 2 shows scree plots for the eigenvalue vectors and their effective ranks as calculated from Equations 3 and 4 (R code is provided in Supplementary Appendix 1 available on Dryad at http://dx.doi.org/10.5061/dryad.d7wm37q01). Effective rank was also calculated for 10,000 matrices in which each column of landmark coordinates was permuted to remove interlandmark covariances. The mean and standard deviations are also reported in Figure 2.

The scree plots of both the correlation and covariance matrices demonstrate that the vector of eigenvalues for each matrix contains 24 nonzero eigenvalues, the expected mathematical rank given the loss of four ranks to the GPA. Eigenvalues 25–28 are nonzero in the permuted matrices showing that this mathematical rank deficiency is recovered when the matrix is randomized. Both real matrices are also highly information rank deficient. The effective rank of the covariance matrix is 11.362, meaning that less than half of the 24 nonzero eigenvalues carry meaningful information. The number of effective ranks in the permuted covariance matrix is about 21, indicating that the covariance matrix is information poor by about seven ranks. Four of these are from the Procrustes analysis; the rest are probably due to insufficient sample size (Grabowski and Porto 2017), as demonstrated by the permuted information content (Fig. 2). The information rank deficiency due to phenotypic covariance is therefore 21 minus the effective rank of the real matrix (11.36), or about 9.64 ranks. Therefore 12.64 ranks, or over half, of the eigenvalue distribution is redundant. The small values of the higher eigenvalues are caused mainly by phenotypic covariance, but they do not carry novel information about it, and lack of matrix information acerbates this deficiency. The eigenvalue spectrum is overdetermined with respect to its information content, and its distribution will be affected as a result. Figure 2 also shows that the rank of the correlation matrix is 12.424. Use of the correlation matrix therefore adds over one full rank of information relative to the covariance matrix. We believe this rank increase is noise, as described in the first section. The pattern for the permuted matrices is similar; the correlation matrix adds almost three ranks of noise over the value for the covariance matrix.

Part 2: Modularity Models and Phenotypic Integration Metrics

In this section, we introduce the Pit 13 dire wolves and show they are more integrated than those from Pit 61/67. We do this by testing both populations for

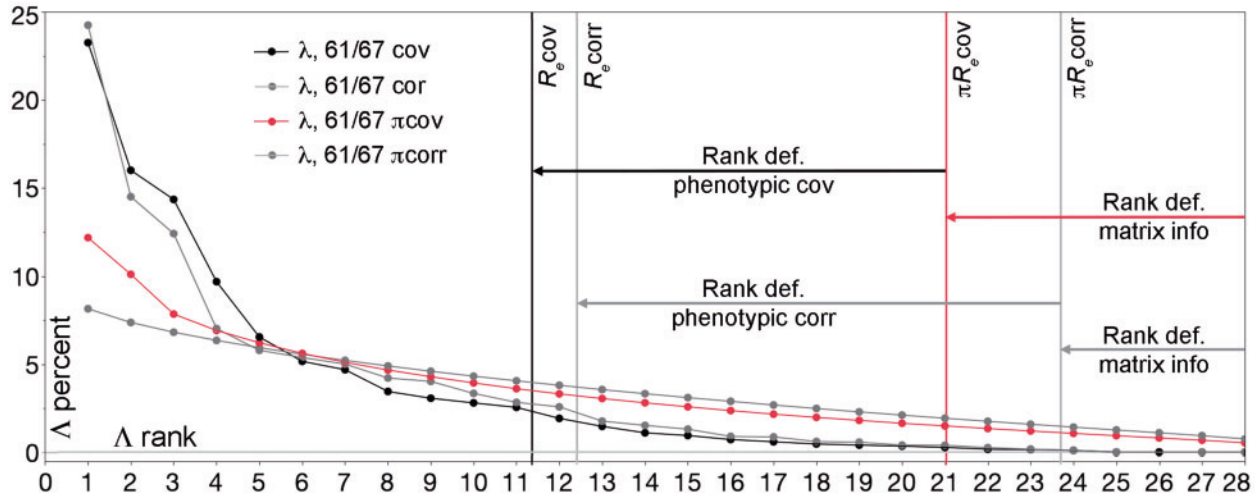

FIGURE 2. Rank deficiency and its sources in Pit 61/67 dire wolves. Scree plots for four data treatments are plotted. Note that the last four eigenvalues are exactly zero for both the correlation and covariance matrices computed from the real data (the mathematical rank = 24), reflecting the loss of four degrees of freedom to the GPA. The effective rank of permuted data shows rank deficiency due to lack of matrix information; the true effective ranks are much smaller than this, and the greater deficiency is due to phenotypic covariance, as well as the four ranks lost to the GPA. Effective ranks for each matrix are: Permuted covariance = 21.028 ± 0.226; permuted correlation = 23.728 ± 0.258; covariance = 11.362; correlation = 12.424.

modularity against a range of candidate two- and three-parameter models. We then calculate the eigenvalue distribution and SVG for the two covariance matrices and show that both fail to capture the modularity change. We then modify the SVG to use only the nonredundant eigenvalues and show that this measure does capture the modularity change.

Modularity models.—A challenge in all modularity studies is initial specification of the models to be tested. In this article, we use a data analytic approach: principal components analysis and multivariate allometry vectors of interlandmark distances are used to identify candidate models. While important, this specification of models is tangential to our discussion of rank deficiency, and we include it below as Supplementary Appendix 2 available on Dryad. This allows us to move directly to testing the candidate models for the two data sets. A total of 14 models of modularity were suggested by data analysis in Supplementary Appendix 2 available on Dryad, and these models were tested individually using the CR statistic (Goswami and Polly 2010; Fig. 3; Table 1), performed in the **geomorph** package in R (Adams et al. 2020). The models were tested both on the pooled data, and on data divided by pit. Effect-size tests and tests against a model of zero modules were performed on 61/67 wolves only, as only these wolves displayed significant modularity; representative tests and results are listed in Table 1.

There are two models with statistically significant CR values (Table 1). Both are two parameter models that contrast the length of the postcanine tooth row with the rest of the jaw. The effect sizes of the modularity models with significant CRs were compared using the **compare.CR** function in **geomorph**; the effect sizes were not significantly different. Comparison with a null model of zero modules was performed using the same function. The cheek teeth are clearly a module for pit 61/67 wolves. For Pit 61/67 wolves only, the cheek tooth model (3–7) was significantly better than the null model, while the model containing only the molars (4–7) was marginally so. The anterior–posterior width of the canine and of the condyloid process (1,2,9,10) do not form a distinct module on their own; the significant CR of the three-module model is driven by the presence of the cheek tooth module. The canine does not take part in this module and therefore has significant variance attributable to another factor. The inability of the modularity tests to identify the canine and condyloid process as a module was surprising. We believe it was not identified because most of its variation covaries with size (Supplementary Appendix 2 available on Dryad). The cheek teeth vary against size, and this allows identification of that module. We note that Pit 13 wolves had no significant modules, and must therefore be more phenotypically integrated than Pit 61/67 wolves.

The magnitude of the CR differences between pits was surprising, given they are taken from populations of the same species at the same location, sampled 5000 years apart. Due to sample size differences (the sample size of Pit 13 was 36 while that of 61/67 was 83), we were concerned that a lack of statistical power was obscuring the modularity signal in Pit 13. To test for this, we jackknifed the 61/67 sample to 36 members to mirror the sample size of Pit 13 before rerunning the CR analysis. Although confidence intervals widened in these subsets, the covariance structure of Pit 61/67 was still evident and significantly modular, indicating that the modularity difference between pits is not attributable to differences in sample size. The pooled CR statistics, run on all 119

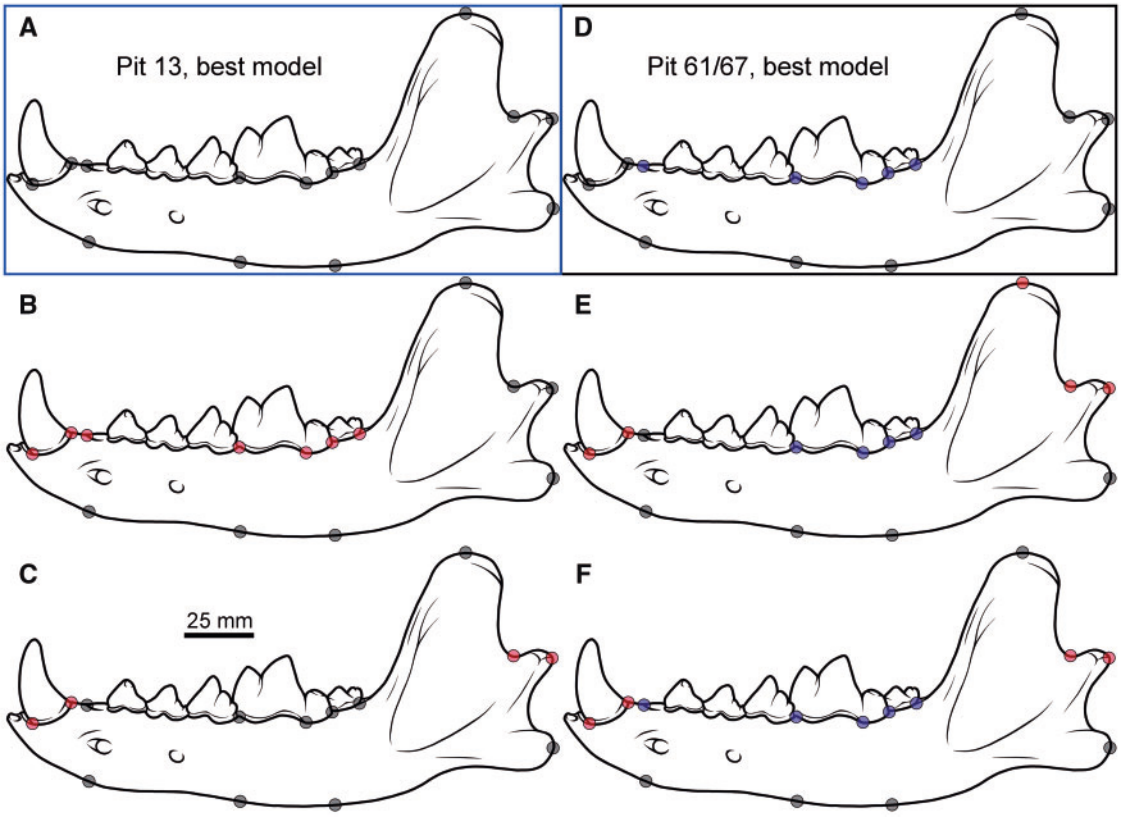

FIGURE 3. Candidate modularity models for the *Canis dirus* jaw tested in this article. The preferred modularity model for Pit 61/67 wolves is D, a two-parameter model of (3,4,5,6,7) versus all other landmarks. This is the model with the best statistical support, but only in Pit 61/67 (Table 1). The preferred model for Pit 13 wolves is A, the one-parameter model. The other models shown did not have significant modularity according to the CR statistic; all models tested are listed in the caption to Table 1.

TABLE 1. Results of representative modularity model tests

Model vs. all other LMs	CR_{pooled}	*P*value	CR_{13}	*P*value	$CR_{61/67}$	*P*value	P_{null}
1,2	1.17	0.427	1.234	0.581	1.063	0.219	—
1,2,9,10	1.10	0.573	1.201	0.838	1.090	0.491	—
1–7	1.15	0.837	1.147	0.786	1.171	0.921	—
3–7	0.96	0.052	1.067	0.387	0.898	0.022*	0.031*
4–7	0.92	0.046*	0.994	0.108	0.894	0.028*	0.046*
1,2,9,10; 3,4,5,6,7	1.06	0.244	1.192	0.719	1.029	0.226	—

Representative module models are listed on the left column, with CR statistics for three data partitions in subsequent columns. Pit 61/67 wolves have two significant modules comprising the check teeth vs. the rest of the jaw, while Pit 13 wolves do not show significant modularity between these or any other modules. As shown in Figure 3, Pit 13 wolves are best fit with a one-parameter model, while Pit 61/67 wolves are best fit with a two-parameter model. Wolves from 61/67 are therefore significantly more modular than those from Pit 13. Other modularity models tested that were not significant include the two-parameter models (4,5,6,7,13,14), (4,5,6,7,8), (4,5,6,7,8,13,14), (1,2,8,9,10), and (1,2,14,9,10,11), and the three-parameter models ((1,2,8,9,10)(4567)), ((1,2,9,10)(4,5,6,7,13,14)), and ((129,10)(4,5,6,7,8)).

wolves, indicate that the inclusion of the Pit 13 wolves actually degrades the global modularity signal (Table 1). This indicates that the relative lack of modularity in Pit 13 is real.

Measures of whole-jaw integration.—The eigenvalue dispersion and SVG metrics were calculated using eigenanalysis, bootstrapping, and permutation codes written in R (R Core Team 2014; Supplementary Appendix 1 available on Dryad). The eigenvalue standard deviation metric is significantly less in Pit 13 wolves (Table 2: $\lambda\sigma$, 13 <61/67, $P = 0.00008$), while the SGV_{24} metric is equivalent between the two groups (13 = 61/67, $P = 0.837$, Student's T in this and subsequent comparisons). This implies that eigenvalue dispersion is significantly less in Pit 13, and those wolves should be more modular than Pit 61/67. Yet this is not the case; the modularity tests reported above clearly show that Pit 61/67 wolves are significantly modular, while Pit 13 wolves are not. The eigenvalue dispersion is less in Pit 13 even though the jaws are more integrated. The SGV_{24} metric fails to detect the increase in modularity.

As demonstrated above, over half of the ranks used to calculate both metrics are redundant, and we believe the SGV_{24} is essentially measuring noise. We experimented with the use of the covariance matrix in the SGV_{24} to see if the failure of that metric was due to use of the correlation matrix. Use of the covariance matrix requires standardization of the eigenvalues before calculation to control for different matrix variances. This modified metric is also reported in Table 2, and like the SGV_{24} it is equivalent between the two matrices. Rank inflation due to use of the correlation matrix is therefore not the reason for failure of the SGV_{24}.

Effective rank-scaled SGV or "effective dispersion".—Calculation of the effective rank for the two populations indicates that the rank of Pit 61/67 is less than that of Pit 13 (Fig. 4: R_e, 13 >61/67, $P = 0.00148$). To account for this difference in dimensionality dispersion, we modify the SGV to consider only the nonredundant eigenvalues of the covariance matrix. This metric includes only the eigenvalues that carry significant information, that is, those up to the effective rank of the matrix. We label this quantity SGV_{Re} and calculate it as follows:

$$SGV_{Re} = \sqrt[Re]{\prod \lambda_{Re}} = \sqrt[Re]{\left(\prod_{i=1}^{[Re]} \lambda_i\right) * (R_e - [R_e])(\lambda_{[Re]+1})}$$

where R_e is the effective rank of the covariance matrix, and $[R_e]$ is the integer value of the effective rank. The SGV_{Re} is the R_e^{th} root of the product of the eigenvalues up to the integer value of R_e, multiplied by the decimal remainder of R_e times the next smallest eigenvalue. The value for this statistic was calculated for the different data partitions using code written in R, with jackknifed confidence intervals for Pit 61/67 (Table 2; Supplementary Appendix 1 available on Dryad). Using this measure, Pit 13 wolves have significantly higher dispersion than those of Pit 61/67. Hence, they are more integrated, while 61/67 wolves are more modular. The SGV_{Re} metric successfully recovers the increase in modularity exhibited by 61/67 wolves.

Previous authors rejected SGV as a statistic to measure dispersion because its distribution has undesirable properties (Pavlicev et al. 2009a); they preferred the standard deviation of the eigenvalues because of its linearity and because it has the same units as the input data. Because the SGV_{Re} is an average variance, one may utilize the definition of the standard deviation to transform the SGV_{Re} into a similar measure:

$$D_e = \sqrt{SGV}_{Re} = \sqrt[2Re]{\left(\prod_{i=1}^{[Re]} \lambda_i\right) * (R_e - [R_e])(\lambda_{[Re]+1})}$$

where the square root of the SGV_{Re} is a new metric, D_e, that we call "effective dispersion." It measures dispersion in both variance and dimensionality together and accounts for the information rank deficiency that misleads the classical integration metrics. The D_e metric is significantly different between the two samples ($P = 0.0049$; Table 2).

Part 3: Matrix Information and Relative Dispersion

Permuted rank standardization.—Sample size impacts effective rank; the effective rank of 61/67 wolves at $n = 36$ is 9.89, while the full matrix of $n = 83$ has an effective rank of 11.362 (Figs. 2 and 4). Clearly the derivation of a version of D_e that accounts for matrix size is desirable, as this will allow comparisons between matrices of different sizes. Pavlicev et al. (2009a) accomplish a similar standardization for eigenvalue standard deviation by dividing the observed eigenvalue standard deviation by its maximum possible value to yield their "relative standard deviation." Because they use the correlation coefficient in their calculations, the minimum possible correlation in a matrix is simply the number of traits minus one, because each trait adds an additional unit of uncorrelated variance. A relative version of effective dispersion is more difficult to calculate because the minimal covariance in a matrix is not derivable from first principles, because the input variance for each variable is not one. Also, because the matrix is information rank deficient, this must also be considered, suggesting that an approach based on effective rank is necessary. The quantity needed for standardization must therefore preserve the variances of the input variables, but remove their statistical covariance, and must also be treated for rank deficiency. This can be construed as a question-specific randomization problem, as described by Manly 1997; Chapter 1). We generate a quantity with the desired properties by permuting the columns in of the landmark data, and then calculating a set of the resulting covariance matrix. These matrices will preserve the input variances on the diagonal, but will remove statistical covariance on the off-diagonal (although covariances will still be nonzero in a rank-deficient matrix). The effective rank of the permuted matrix can then be calculated, yielding πR_e.This permuted effective rank is the maximum matrix rank given the variable input variances, and no statistical correlation among them. It is the amount of matrix information and is driven by sample size in reasonably complex shapes. It can be used as a basis for standardizing D_e for matrix size. The effective dispersion is the mean eigenvalue variance limited to its nonredundant information content, and we wish to scale D_e to an expectation of maximum possible information content per axis. We are not concerned with scaling for the variance in the matrix, but for the amount of information in one dimension of an uncorrelated matrix that conserves this variance. The appropriate scaling term is therefore one rank of the permuted matrix, or $1/\pi R_e$. This yields the definition for the "relative dispersion," or D_r:

$$D_r = \frac{D_e}{\sqrt{\frac{1}{\pi R_e}}} = \frac{\sqrt[2Re]{\prod \lambda_{Re}}}{\sqrt{\frac{1}{\pi R_e}}}$$

TABLE 2. Whole jaw integration measures for the *Canis dirus* mandible data sets

Data partition	$\lambda\sigma, \pm$	$SGV_{24}, \pm$	$covSGV_{24}, \pm$	$SGV_{Re}, \pm$	$D_e \pm$
Canis dirus 13	0.277	0.363	0.0108	4.131e−5	0.0064
$n=36$	—	—	—	—	—
Canis dirus 6167j	0.304	0.362	0.0108	3.496 e−5	0.0058
$n=36$	0.027	0.029	0.0008	1.42 e−5	0.0012

All reported confidence intervals are one standard deviation of 10,000 jackknife or permutation replicates. The D_e metric is significantly different between the two samples ($P = 0.0049$). Quantities listed are eigenvalue standard deviation (correlation matrix, 24 eigenvalues), ($\lambda\ \sigma$); standardized generalized variance 24 (**SGV_{24}**); standardized generalized variance 24 using the covariance matrix, and standardizing the eigenvalue vector to the sum of eigenvalues (**$covSGV_{24}$**); standardized generalized variance effective rank (**SGV_{Re}**); and effective dispersion (**D_e**).

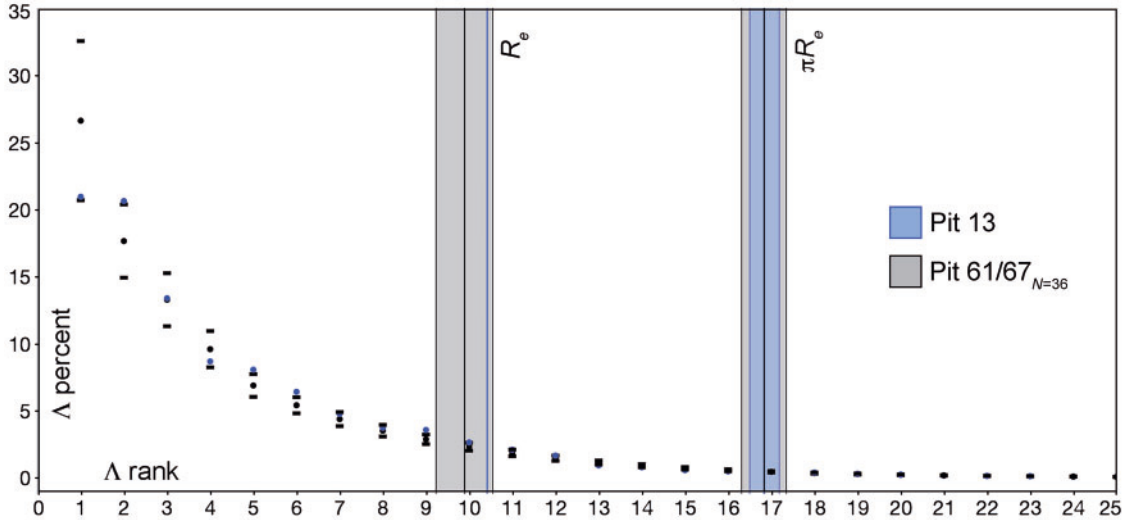

FIGURE 4. Scree plots and effective rank for GPA landmark data of *Canis dirus* jaws. Blue is Pit 13, black is Pit 61/67. Confidence intervals for effective rank (R_e) are 10,000 jackknife replicates of the Pit 61/67 sample to an $n=36$. Eigenvalue dispersion appears greater in Pit 61/67 wolves, even though they have significant modularity and Pit 13 wolves do not. However the amount of total variance is greater in Pit 13, and the Pit 13 effective rank is significantly larger, so the variance is spread over a greater number of dimensions. The permuted dimensionality is also shown on both plots; this value is identical between pits ($\pi R_e = 16.84$), demonstrating that the rank deficiency difference between the pits is due to phenotypic covariance only (i.e., the rank deficiency due to lack of matrix information is equivalent). The effective rank of both matrices is very deficient, with fewer than half of the 24 eigenvalues being meaningful. The effective rank of Pit 13 wolves is 10.414; the permuted effective rank is 16.841 ± 0.345. The jackknifed effective rank for Pit 61/67 wolves is 9.89 ± 0.656, while the permuted value at $n=36$ is 16.832 ± 0.519.

Relative Dispersion of Canis dirus *and* Smilodon.—To demonstrate the use of relative dispersion, we turn to the third data set used in this article, that of RLB *Smilodon fatalis* (Meachen et al. 2014). These data comprise 14 2D landmarks on 81 jaws, hence, are quite similar to the dire wolf data (11 of the 14 landmarks are homologous between *Smilodon* and *Canis dirus*). We began by calculating the effective rank for *Smilodon*, and comparing it to Pit 61/67 dire wolves (Fig. 5). Both samples were bootstrapped for confidence intervals, and the sample size for 61/67 was held at 81 during the bootstrap. Both resampling procedures lower effective sample size, and so reduce matrix information and hence effective rank. The exact values for integration statistics are reported in Table 3, along with the bootstrap means and standard deviations. Figure 5 plots the $n=81$ bootstrap values for effective rank.

Smilodon and *Canis dirus* jaws are similar in that covariance matrices for both are highly overdetermined with respect to their effective rank, although the effective rank of *Smilodon* is significantly greater. The permuted ranks of both taxa are much closer to 24 at a sample size of 81 compared to *Canis dirus* Pit 13, at a sample size of 36. This implies that the permuted rank of the matrices should converge toward the expected mathematical rank of 24 at large n (probably over 100, in accord with sample size requirements of other integration metrics as shown by Grabowski and Porto 2017). The relative dispersion is comparable among data sets of different sample sizes, as demonstrated by the full and jackknifed 61/67 sample; the D_e of the 61/67 $n=36$ jackknife sample is 0.0058, while that of the full $n=83$ data set is 0.0052; the D_r of both partitions is equivalent, at 0.0238). This, and the values in Table 3, demonstrates that the relative dispersion functions as intended; it is robust to difference in sample size, and successfully captures the modularity evolution between dire wolf populations. The comparison for D_r is marginally significant, while that for sample-size controlled D_e was strongly significant (Table 2). Lastly, the D_r metric shows that the *Smilodon* jaw is much more tightly integrated than that of *Canis dirus*. We note that

TABLE 3. Whole jaw integration measures for the *Canis dirus* jaw samples, and for the *Smilodon* data

Data partition	$\lambda\sigma_{24}$, ±	SGV_{24}	SGV_{Re}	D_e	D_r
Canis dirus 13 $n=36$	0.277	0.363	4.131e−5	0.0064	0.0263
BS estimate	0.316	—	5.029e−5	0.00690	0.0283
BS sd	0.015	—	2.283e−5	0.00162	0.0066
Canis dirus 6167 $n=83$	0.278	0.500	2.71e−5	0.0052	0.0238
BS estimate	0.296	0.421	3.373e−5	0.00567	0.0260
BS sd	0.025	0.030	1.311e−5	0.00112	0.0051
Smilodon $n=81$	0.250	0.575	7.20e−5	0.0085	0.0391
BS estimate	0.270	0.480	9.098e−5	0.00940	0.0432
BS sd	0.010	0.024	3.055e−5	0.00161	0.0074

The relative dispersion difference between Pit 13 and Pit 61/67 wolves is not quite significant (Welch's $T=1.863$, $P=0.068$); Those between *Smilodon* and the two wolf data sets are highly significant (13, $T=-10.85$, $P<0.0001$; 6167, $T=-17.29$, $P<0.0001$). The *Smilodon* jaw is much more tightly integrated than that of *Canis dirus*. Note that the SGV_{24} also correctly captures the integration difference between *Smilodon* and *Canis dirus* while the eigenvalue standard deviation still fails. Quantities listed are eigenvalue standard deviation (correlation matrix, 24 eigenvalues), (λ σ); standardized generalized variance 24 (**SGV_{24}**); standardized generalized variance effective rank (**SGV_{Re}**); effective dispersion (**D_e**); and relative dispersion (**D_r**).

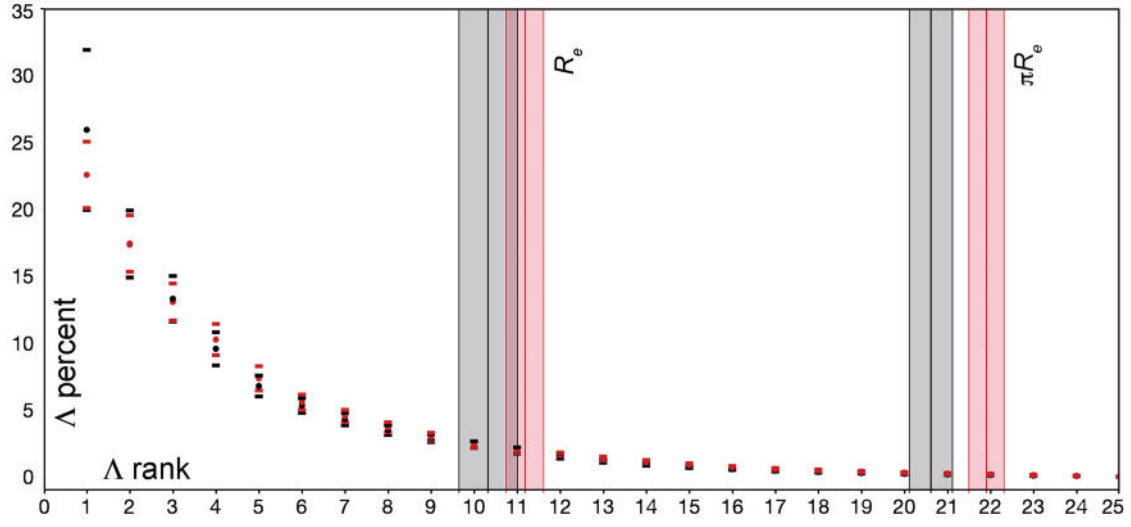

FIGURE 5. Scree plot comparison for *Canis dirus* and *Smilodon* data sets, 14 landmarks on jaws, at $n=81$. Plotting conventions are the same as those used in Figure 2. *Canis dirus* 61/67 is shown in black, *Smilodon* in red. Both data sets have been bootstrapped for a confidence interval (10,000 replicates), hence the values shown here are less than the precise values of R_e for *C. dirus* (11.36) and *Smilodon* (12.32; Table 3). Because bootstrapping lowers the effective sample size, this drop in effective rank illustrates the strong sensitivity of effective rank to sample size. The plotted means and standard deviations are: *Canis dirus* 61/67, $R_e = 10.33 \pm 0.677$; $\pi R_e = 20.614 \pm 0.503$. *Smilodon* $R_e = 11.184 \pm 0.438$; $\pi R_e = 21.91 \pm 0.503$.

the SGV_{24} successfully recovers this difference, while the eigenvalue standard deviation is still incorrect.

DISCUSSION

Sources of Rank Deficiency and Its Treatment

In this paper we demonstrate that a typical Procrustes-superimposed landmark covariance matrix is highly rank deficient with respect to its information content. This is evident from the scree plots, which show many small eigenvalues. Quantifying this information rank deficiency requires use of the Shannon, or information, entropy. We operationalize the Shannon entropy in a geometric morphometric context and use it to calculate that the information rank, or effective rank, of the dire wolf covariance matrix is less than half of the expected mathematical rank. Geometric morphometric data sets have several characteristics that make effective rank deficiency more severe than in matrices of linear measures. The first is the GPA procedure, which utilizes four degrees of freedom and hence removes four ranks from the covariance matrices dealt with here. The magnitude of this rank loss is known (Fig. 1) and can be accommodated. By contrast, the magnitude of effective rank deficiency due to lack of matrix information is not known. We use a permutation approach to estimate this magnitude, allowing us to calculate the magnitudes of effective rank deficiency due to phenotypic covariance versus the other sources. Phenotypic covariance accounts for about 9.6 ranks of deficiency in the dire wolf data set considered in Part 1, while lack of matrix information accounts for about 3 ranks, and the GPA accounts for 4 ranks. This nonphenotypic rank deficiency is worrisome, because

it misleads phenotypic integration metrics that are calculated from the entire eigenvalue spectrum.

To demonstrate the impact of effective rank deficiency on current integration metrics, we show that current phenotypic integration metrics fail to identify the modularity change between dire wolf populations. Granted, this is a minor allometric change, from a one-parameter model to a two-parameter model, and the differences in magnitudes are not large. However, eigenvalue standard deviation is statistically significant in the wrong direction, while the SGV_{24} lacks sensitivity. The SGV_{24} does successfully capture the modularity difference between *Smilodon* and *Canis dirus*, where the integration difference is large. Yet eigenvalue standard deviation continues to fail. We also show that use of the correlation matrix leads to effective rank inflation in coordinate data; this rank inflation is noise, and the correlation matrix should not be used on raw coordinate data.

The failure of current integration metrics arises from the inclusion of the long tail of small, nonzero eigenvalues in the eigenvalue distribution. This tail of small values will have low dispersion, and eigenvalue standard deviation will be pushed lower as effective rank deficiency increases. The geometric mean of the eigenvalue distribution (the SGV_{24}) will also be pushed lower, again due to the inclusion of many small eigenvalues. These facts are true on inspection but should be demonstrated; one immediate line of future research is a simulation study that quantifies the sensitivity of current phenotypic integration metrics to information rank deficiency. This article is meant as a theoretical development and proof of concept, and we do not attempt a simulation study here, although the behavior of the effective and relative dispersion metrics should also be explored by simulation. The seeming loss of statistical power between D_e and D_r deserves specific attention.

The use of the effective rank allows calculation of a modified form of the geometric mean of the eigenvalues, which is simply the mean of the nonredundant eigenvalues. We demonstrate that this new metric, D_e, successfully recovers the evolution of increased modularity in Pit 61/67 wolves. However, matrix effective rank is very sensitive to sample size (Table 3), and we held sample size constant in Part 2. Therefore, we introduce the concept of relative dispersion, D_r, to account for differences in matrix information content in Part 3. Relative dispersion accounts for dispersion in dimensionality, as well as dispersion in variance, and it accounts for differences in matrix size by standardizing against the permuted information content. The metric D_r should therefore be comparable among data sets, as we demonstrate via comparison with *Smilodon*. This comparison utilizes data sets with an equal number of landmarks (14); the sensitivity of relative dispersion to different landmark number requires quantification in a manner similar to the analysis of classical integration metrics by Grabowski and Porto (2017).

Dense Semilandmarks and Latent Dispersion

The relative dispersion metric accounts for information rank deficiency, but it is still space-specific. By this, we mean that it is calculated in a shape space defined by the landmarks, not the entire shape, and the relation between these two spaces is not clear. For maximum utility, it is desirable that D_r should be comparable among spaces defined by different, arbitrary sets of landmarks. In the case of two spaces defined by the same number of landmarks, as in the *Canis dirus* and *Smilodon* data compared here, one might suppose that relative dispersion is comparable even if the landmarks are not homologous, because the full and mathematical ranks are the same (11 of the 14 landmarks are in fact homologous). However, the effective ranks are not the same, and so the spaces are not comparable. A landmark-defined space with covariance matrix $\mathbf{K_V}$ and an effective rank based on the Shannon entropy only captures the variance of the landmarks, not of the entire shape upon which those landmarks occur. One way to represent the total covariance of the entire shape, $\mathbf{K_T}$, would be to sample the shape with an arbitrarily large number of landmarks, so that as the magnitude of $\mathbf{V}$ increases $\mathbf{K_V}$ would approach $\mathbf{K_T}$. Thought of in this way, a typical set of geometric morphometric data would comprise

$$K_T = K_V + K_R$$

where $\mathbf{K_R}$ is the residual covariance in the shape not captured by the landmarks in $\mathbf{K_V}$. The residual matrix $\mathbf{K_R}$ is of unknown magnitude in all mainstream applications of geometric morphometrics, and there is no guarantee that it is negligible, nor that the proportion $\mathbf{K_R/K_T}$ is of constant magnitude among spaces, even if $\mathbf{K_V}$s are of equivalent effective rank. Therefore, the relative dispersion metric defined above remains space-specific.

A recent approach to estimating residual shape variance is the employment of dense semilandmarks, where a curve or surface is first landmarked, and then densely sampled with an increasing number of semilandmarks until the coefficient of interest stabilizes (Marshall et al. 2019). This procedure can be employed here. Given a shape with variance $\mathbf{K_T}$ and landmark variance $\mathbf{K_V}$, the residual covariance $\mathbf{K_R}$ will decrease as $\mathbf{V}$ increases, so that

$$K_T \approx K_V, \text{ and } K_R \approx 0$$

as the number of semilandmarks becomes large. The number of semilandmarks can be arbitrarily large, down to the pixel or voxel resolution of the image being digitized, but in practice semilandmarks need only be added until the coefficient stabilizes. Because the relative dispersion metric relies on information entropy, it will scale with the information added by each landmark, not with landmark number. Using this procedure allows characterization of the true, or latent, dispersion of the shape;

$$D_l = \lim_{v \to \infty} D_r$$

where the latent dispersion of the shape space, D_l, is the relative dispersion of a matrix that is semilandmarked densely enough to asymptotically stabilize D_r. Classical phenotypic integration measures are not amenable to this procedure, because they consider all eigenvalues, and as the number of landmarks increases the mathematical rank of **K** will increase without bound. We contend that the effective rank will converge as landmark number increases; developing and implementing methods to test this hypothesis is a topic of current research.

If it does converge, the latent dispersion metric should be useful for comparing shapes across arbitrary landmark spaces, and for assessing the fidelity with which the original landmark data capture phenotypic shape change. It is a general statement of the information content of a shape. Calculation of this metric would allow us to state definitively that the *Smilodon* data are tighter in shape dispersion than *Canis dirus*, assuming the observed pattern from 14 landmarks holds up. Yet given the effective rank deficiency in **K** demonstrated by the permuted matrices, we doubt there is sufficient residual variation not being captured by the landmarks to substantially change this result. *Smilodon* is a highly specialized hypercarnivore, and intuitively it should be more tightly adapted—and less complex—than a more generalized canid like *Canis dirus*, whose jaw should be more information rich.

Conclusions

This article reviews two current measures of phenotypic integration, and how they interface with modularity measures. VanValen's realization that dimensionality dispersion is a critical property is highlighted, and the Shannon, or information, entropy is employed to measure it as effective rank. Phenotypic covariance matrices for dire wolf jaws from two populations are then introduced: Pit 13, deposited circa 19 kya, and Pit 61/67, deposited circa 14 kya. Both data matrices comprise coordinate data for 14 identical landmarks, after Procrustes superimposition. Use of the concept of information entropy allows identification of the magnitude and sources of rank deficiency in the Pit 61/67 sample. Effective rank deficiency is found to be high, with almost half not due to phenotypic covariance. Modularity model tests show that the only significant module is that of the cheek teeth relative to the jaw corpus. This module is only significant in Pit 61/67 wolves; its absence in Pit 13 means that Pit 13 wolves are more integrated (best fit by a one parameter model), while 61/67 wolves are more modular (best fit by a two parameter model). Classical metrics of phenotypic integration fail to recover the evolutionary increase in modularity in Pit 61/67 wolves. They fail because i) they rely on the correlation matrix, which inflates variance in geometric morphometric data; and ii) they measure only variance dispersion, while ignoring dimensionality dispersion.

New metrics based on the Shannon entropy-modified SGV are defined to quantify the dispersion of phenotypic shape spaces. The relative dispersion (D_r) is a sample-size corrected version of the effective dispersion (D_e); the latent dispersion is an asymptotic extension of D_r to a dense semilandmark context. The relative dispersion is the core result of this paper. It is a phenotypic integration measure that has been normalized to the information content of the eigenvalue spectrum, and we restate it from Equation 7:

$$D_r = \frac{\sqrt[2Re]{\prod \lambda_{Re}}}{\sqrt{\frac{1}{\pi R_e}}}.$$

New whole phenotypic shape integration measures incorporating effective rank are successful at recovering the evolution of increased modularity in 61/67 wolves and show promise for the characterization of phenotypic spaces among taxa.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.d7wm37q01.

Acknowledgements

We thank the curators and staff at the Tar Pit Museum, particularly John Harris, Emily Lindsey, Chris Shaw, Aisling Farrell, and Gary Takeuchi, for allowing us to collect data from the dire wolf specimens used in this project and for assistance of all kinds. We thank the Marshall University Drinko Research Fellowship to FRO for funding towards this project. Special thanks are due to Emily Lindsey, John Southon, Wendy Binder, and Larisa DeSantis for input on early versions of this manuscript, and to Bryan Carstens, Lauren Esposito, and two anonymous reviewers for thorough and very helpful input during the review process. Thanks are due to Thom Yorke and Noel Gallagher. Work by FRO was funded in part by NSF EAR-SGP 1757236 (Project SABER) and by a Drinko Distinguished Research Fellowship to F.R.O.

References

Adams D.C. 2016. Evaluating modularity in morphometric data: challenges with the RV coefficient and a new test measure. Methods Ecol. Evol. 7:565–572.

Adams D., Collyer M., Kaliontzopoulou A. 2020. Geomorph: Software for geometric morphometric analyses. R package version 3.2.1. https://cran.r-project.org/package=geomorph.

Binder W.J., Thompson E.N., Van Valkenburgh B. 2002. Temporal variation in tooth fracture among Rancho La Brea dire wolves. J. Vertebrate Paleontol. 22:423–428.

Binder W.J., Van Valkenburgh B. 2010. A comparison of tooth wear and breakage in Rancho La Brea sabertooth cats and dire wolves across time. J. Vertebrate Paleontol. 30:255–261.

Bookstein F.L. 1997. Morphometric tools for landmark data: geometry and biology. Cambridge, UK: Cambridge University Press.

Bookstein F.L. 2015. Integration, disintegration, and self-similarity: characterizing the scales of shape variation in landmark data. Evol. Biol. 42:395–426.
Brannick A.L., Meachen J.A., O'Keefe F.R. 2015. Microevolution of jaw shape in the dire wolf, Canis dirus, at Rancho La Brea. In: Harris J.M., editor. La Brea and beyond: the paleontology of asphalt-preserved biotas. Los Angeles: Natural History Museum of Los Angeles County, Science Series, vol. 42. p. 23–32.
Bunea F., She Y., Wegkamp M.H. 2011. Optimal selection of reduced rank estimators of high-dimensional matrices. Ann. Stat. 39:1282–1309.
Cangelosi R., Goriely A. 2007. Component retention in principal component analysis with application to cDNA microarray data. Biol. Direct 2:1–21.
Cheverud J.M. 1982. Phenotypic, genetic, and environmental morphological integration in the cranium. Evolution 36:499–516.
Cheverud J.M. 1996. Developmental integration and the evolution of pleiotropy. Am. Zool. 36:44–50.
Cheverud J.M., Rutledge J.J., Atchley W.R. 1983. Quantitative genetics of development: genetic correlations among age-specific trait values and the evolution of ontogeny. Evolution 895–905.
Curth S., Fischer M.S., Kupczik K. 2017. Patterns of integration in the canine skull: an inside view into the relationship of the skull modules of domestic dogs and wolves. Zoology 125:1–9.
Darwin C. 1859. On the origin of species by means of natural selection. London, UK: John Murrary.
Drake A.G., Klingenberg C.P. 2010. Large-scale diversification of skull shape in domestic dogs: disparity and modularity. Am. Nat. 175:289–301.
Dryden I.L., Mardia K.V. 1998. Statistical shape analysis. Chichester, UK: Wiley. 347 p.
Grabowski M., Porto A. 2017. How many more? Sample size determination in studies of morphological integration and evolvability. Methods Ecol. Evol. 8:592–603.
Goswami A., Smaers J.B., Soligo C., Polly P.D. 2014. The macroevolutionary consequences of phenotypic integration: from development to deep time. Philos. Trans. R. Soc. B 369(1649):20130254.
Goswami A., Binder W.J., Meachen J., O'Keefe F.R. 2015. The fossil record of phenotypic integration and modularity: a deep-time perspective on developmental and evolutionary dynamics. Proc. Natl. Acad. Sci. USA 112:4891–4896.
Goswami A., Polly P.D. 2010. Methods for studying morphological integration and modularity. Paleontol. Soc. Pap. 16:213–243.
Hine E., Blows M.W. 2006. Determining the effective dimensionality of the genetic variance–covariance matrix. Genetics 173:1135–1144.
Horvath S. 2011. Weighted network analysis: applications in genomics and systems biology. Berlin, Germany: Springer Science & Business Media.
Jolicoeur P. 1963. 193. Note: the multivariate generalization of the allometry equation. Biometrics 19:497–499.
Klingenberg C.P. 1996. Multivariate allometry. In: Marcus L.F., Corti M., Loy A., Naylor G.J.P., Slice D.E., editors. Advances in morphometrics. Boston, MA: Springer. p. 23–49.
Klingenberg C.P. 2008. Morphological integration and developmental modularity. Annu. Rev. Ecol. Evol. Syst. 39:115–132.
Klingenberg C.P. 2010. Evolution and development of shape: integrating quantitative approaches. Nat. Rev. Genet. 11:623–635.
Klingenberg C.P. 2013. Cranial integration and modularity: insights into evolution and development from morphometric data. Hystrix 24:43–58.
Klingenberg C.P., Zaklan S.D. 2000. Morphological integration between developmental compartments in the Drosophila wing. Evolution 54:1273–1285.
Manly B.F.J. 1997. Randomization, bootstrap, and Monte Carlo methods in biology. 2nd ed. Texts in statistical science. UK: Chapman and Hall.
Marshall A.F., Bardua C., Gower D.J., Wilkinson M., Sherratt E., Goswami A. 2019. High-density three-dimensional morphometric analyses support conserved static (intraspecific) modularity in caecilian (Amphibia: Gymnophiona) crania. Biol. J. Linnean Soc. 126:721–742.
Meachen J.A., O'Keefe F.R., Sadleir R.W. 2014. Evolution in the sabre-tooth cat, *Smilodon fatalis*, in response to Pleistocene climate change. J. Evol. Biol. 27:14–723.
Meloro C. 2012. Mandibular shape correlates of tooth fracture in extant Carnivora: implications to inferring feeding behavior of Pleistocene predators. Biol. J. Linnean Soc. 106:70–80.
Najarzadeh D. 2019. Testing equality of standardized generalized variances of k multivariate normal populations with arbitrary dimensions. Stat. Methods Appl. 28:593–623.
O'Keefe F.R., Rieppel O., Sander P.M. 1999. Shape disassociation and inferred heterochrony in a clade of pachypleurosaurs (Reptilia, Sauropterygia). Paleobiology 25:504–517.
O'Keefe F.R., Binder W.J., Frost S.R., Sadlier R.W., Van Valkenburgh B. 2014. Cranial morphometrics of the dire wolf, *Canis dirus*, at Rancho La Brea: temporal variability and its links to nutrient stress and climate. Palaeontol. Electron. doi: 10.26879/437.
O'Keefe F.R., Meachen J., Fet E.V., Brannick A. 2016. Ecological determinants of clinal morphological variation in the cranium of the North American gray wolf. J. Mammal. 94: 1223–1236.
Olson E.C., Miller R.L. 1958. Morphological integration. Chicago: University of Chicago Press (reprint, 1999).
Pavlicev M., Cheverud J.M., Wagner G.P. 2009a. Measuring morphological integration using eigenvalue variance. Evol. Biol. 36:157–170.
Pavlicev M., Wagner, G.P., Cheverud J.M. 2009b. Measuring evolutionary constraints through the dimensionality of the phenotype: adjusted bootstrap method to estimate rank of phenotypic covariance matrices. Evol. Biol. 36:339–353.
R Core Team. 2014. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: https://www.R-project.org/.
Rohlf F.J. 1990. Rotational fit Procrustes methods. University of Michigan Museum of Zoology Special Publications, 2:227–236.
Roy O., Vetterli M. 2007. The effective rank: a measure of effective dimensionality. In 2007 15th European Signal Processing Conference. IEEE. p. 606–610.
Segura V., Cassini G.H., Prevosti F.J., Machado F.A. 2021. Integration or modularity in the mandible of Canids (Carnivora: Canidae): a geometric morphometric approach. J. Mammal. 28:145–157.
SenGupta A. 1987. Tests for standardized generalized variances of multivariate normal populations of possibly different dimensions. J. Multivariate Anal. 23:209–219.
Shannon C.E. 1948. A mathematical theory of communication. Bell Syst. Technical J. 27:379–423.
Stock C., Harris J.M. 1992. Rancho La Brea: a record of Pleistocene life in California. Science Series No. 37. Los Angeles, CA: Natural History Museum of Los Angeles County.
Van Valen L. 1974. Multivariate structural statistics in natural history. J. Theor. Biol. 45:235–247.
Van Valkenburgh B. 1988. Incidence of tooth fracture among large, predatory mammals. American Naturalist 13:291–302.
Van Valkenburgh B. 2009. Costs of carnivory: tooth fracture is Pleistocene and recent carnivores. Biol. J. Linnean Soc. 96:68–81.
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.
Wilks S.S. 1932. Certain generalizations in the analysis of variance. Biometrika 24:471–494.
Zelditch M.L., Bookstein F.L., Lundrigan B.L. 1992. Ontogeny of integrated skull growth in the cotton rat *Sigmodon fulviventer*. Evolution 46:1164–1180.
Zelditch M.L., Swiderski D.L., Sheets H.D. 2012. Geometric morphometrics for biologists: a primer. Cambridge: Academic Press.