Shape Dimensionality Metrics for Landmark Data

F. Robin O'Keefe[1*]

Marshall University, Biological Sciences, Huntington, WV

*corresponding author: F. Robin O'Keefe, Professor, Marshall University, College of Science 265, One John Marshall Drive, Hunington, WV 25755. Phone: +1 304 696 2427. Email: okeefef@marshall.edu

Running Head: Whole Shape Integration Metrics

Keywords: Dire wolf, *Canis dirus*, geometric morphometrics, modularity and integration, information entropy, effective rank, effective dispersion, latent dispersion.

ABSTRACT

The primary goal of this paper is to examine and rationalize different integration metrics used in geometric morphometrics, in an attempt to arrive at a common basis for the characterization of phenotypic covariance in landmark data. We begin with a model system; two populations of Pleistocene dire wolves from Rancho La Brea that we examine from a data-analytic perspective to produce candidate models of integration. We then test these integration models using the appropriate statistics and extend this characterization to measures of whole-shape integration. We demonstrate that current measures of whole-shape integration fail to capture differences in the strength and pattern of integration. We trace this failure to the fact that current whole-shape integration metrics purport to measure only the pattern of inter-trait covariance, while ignoring the dimensionality across which trait variance is distributed. We suggest a modification to current metrics based on consideration of the Shannon, or information, entropy, and demonstrate that this metric successfully describes differences in whole shape integration patterns. Finally, the information entropy approach allows comparison of whole shape integration in a dense semilandmark environments, and we demonstrate that the metric introduced here allows comparison of shape spaces that differ arbitrarily in their dimensionality and landmark membership.

The study of integration in biological systems has a long history, stretching back to Darwin (1859), given a modern footing by Olson and Miller (1958), quantified by Cheverud (1982, 1996), and maturing into a broad topic of modern inquiry encompassing genetics, development, and the phenotype (see Klingenberg, 2013, and Goswami and Polly, 2010, for reviews). The concept of 'integration' means that the traits of a biological whole are tightly dependent. This dependence 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, Figure 1). Yet trait dependency can also remove constraint by giving a population access to novel areas of adaptive space (Goswami et al., 2014, Figure 5). Consequently the integration of traits is central to the evolvability of biological systems, accounting for the intense research scrutiny it has received. This work has produced a family of metrics that summarize and characterize covariance patterns in quantitative morphological traits (Pavlicev et al, 2009; Goswami and Polly, 2010), each with its own mathematical treatment and simplifying assumptions.

Background

"The diagonalization of a matrix is the way in which the space described by the matrix can be conveniently summarized, and the properties of the matrix determined."—Blows, 2006, p.2.

The use of landmark data to characterize the shape of biological structure, and to quantify shape change, has become ubiquitous since the introduction of geometric morphometrics by Bookstein (1997; reviewed in Klingenberg, 2010). All geometric morphometric

techniques begin with a matrix **LM** of landmark data comprising spatial coordinates of homologous points on a series of specimens, that may be of two or three dimensions, so that the elements of the variable vector **V** are:

$$v_i = lm_i(x_i, y_i), \text{or } lm_i(x_i, y_i, z_i)$$

for $\mathbf{LM_{n,2i}}$ or $\mathbf{LM_{n,3i}}$. These landmark positions are measured on *n* samples from the groups of interest, subject to the stricture that each specimen has all of the landmarks. While methods exist to impute missing data, inclusion of a subset of specimens that lack a homologous landmark entirely is not tractable methodologically. Additionally, as shapes become more different, the intersection of their landmark spaces declines, as does the utility of the resulting subspace. Each landmark shape space is therefore unique, and while each space is useful in itself, generalizing among shape spaces is problematic.

Analysis of geometric morphometric landmark data generally begins with Generalized Procrustes Analysis (GPA) of the coordinate data, wherein the centroid of each specimen is translated to the origin, and the coordinates are rotated and scaled to a mean shape so that the intra-landmark variance among specimens is minimized (Bookstein 1997). Given a matrix $\mathbf{LM_{n,v}}$ of *v* landmark coordinates and *n* specimens, a matrix resulting from Procrustes superimposition is:

$$\mathcal{X}_{n,v} = GPA(LM_{n,v})$$

The matrix **X** contains the GPA-transformed landmark coordinates and is of the same order as **LM**. This procedure also produces a vector of centroid size. This vector is interesting in itself, and can also be used to remove size-correlated shape variation from **X** by taking the residuals of its regression against centroid size (Drake and Klingenberg 2010). The desirability of removing size-correlated shape variation prior to subsequent analysis is question specific, and is discussed further below (see Klingenberg, 2010, and Zelditch et al., 2012 for reviews of GPA).

Most subsequent analyses of **X** proceed via Principal Components Analysis (PCA), a factor-analytic expression of the singular value decomposition (SVD), performed on the covariance matrix **K** computed from the variables in **X**;

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

The matrix **K** is square and symmetric with covariances on the off-diagonal and the variances of v_i on the diagonal. The full rank of **K** is v; however, the GPA procedure utilizes four degrees of freedom present in the original data, so the full rank of **K** is v - 4 (Zelditch et al., 2012). The SVD produces three matrices, one of which is the new covariance matrix **Λ**, that is equivalent to **K**, but contains zeros on the off-diagonal and the variances of eigenvectors on the diagonal:

$$\Lambda = SVD(K)$$

These variances are termed the eigenvalues, and each is associated with an eigenvector that is a linear combination of the original variables (Blows, 2006). The eigenvectors are mutually orthogonal, and the eigenvalues descend in magnitude, so that the first eigenvalue is the largest, followed by the second, etc. The vector of eigenvalues $\mathbf{\Lambda_v}$ can be extracted from $\mathbf{\Lambda}$ for further study, and used in its raw form, or standardized by division by the trace of $\mathbf{\Lambda}$ to create a vector expressing the proportion of total variance explained by each component, and plotted as the familiar scree plot.

The interlandmark distances computed from **LM**, or **X**, have utility, but these are seldom utilized on their own (but see Cheverud, 1982; Lele and Richtsmeier, 1991). Geometric morphometric data has the complication that each landmark is represented by two or three coordinates in **X**. There are two ways to handle this; one can either concatenate the landmark coordinates into a single measure before analysis (Goswami and Polly, 2010, use the congruence coefficient), or the coordinates can be analyzed as if they were independent variables. There are benefits and drawbacks to both approaches. Use of the congruence coefficient has been criticized by Klingenberg (2008), who claimed it does not reliably capture correlations among landmarks that covary in different directions. However the use of raw coordinates treats each coordinate as independent within a single landmark, when they are not. A correlation matrix derived directly from **X** will therefore have artificially high full rank. Individual coordinates pose a further problem when a correlation matrix is used, because landmark displacements that are dominated by a vector parallel to one axis will have trivial variation on other axes, and in a correlation matrix this trivial variation will be awarded full rank and expanded to a variance of one. A correlation matrix of **X** will therefore be overdetermined, and possess

a degree of noise from inflation of trivial variance (for discussion of the desirability of the covariance matrix in morphometric applications see Goswami and Polly, 2010, p. 217). Therefore use of the correlation matrix is unwise with morphometric data unless an approach like the congruence coefficient is used. All novel calculations in this paper use the covariance matrix of **X** because this matrix preserves all variation for each landmark, while preserving its correct magnitude. This matrix should be overdefined, and we demonstrate that it is below. For classical measures of whole shape integration we use the correlation matrix, as stipulated in the initial definition of these metrics.

There are myriad examples in the literature of the application of principal components analysis to GPA landmark data (for example, Segura et al., 2020, for canid dentaries; Brannick et al., 2015, for dire wolf dentaries, and O'Keefe et al. 2014, for dire wolf crania). The usual analysis progression is to calculate the principal component scores for the first few PCs, and plot the specimens into the principal component space. Factor analytic interpretation of shape differences along PC axes is complicated compared to traditional morphometrics (Marcus 1990) because the eigenvector coefficients are not interpretable as 'loadings' in the factor-analytic sense (Klingenberg, 2010). However, various methods exist to visualize the landmark displacements along PCs, allowing biological interpretation of axes. Principal components analysis is therefore a powerful tool for dimensionality reduction, allowing the visualization and biological interpretation of a large proportion of the variance in **K** on one or a few axes, and this has been its primary use in geometric morphometrics (Zelditch et al 2012).

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.

If integration is related to covariance between traits, then the degree to which the traits are independent is an indicator of integration. The rank of the covariance or correlation matrix and the distribution of its eigenvalues have been used in several ways by previous authors to derive metrics of integration. Matrix rank is the dimension of the vector space spanned by the variables of the matrix. If all traits are completely independent then the matrix has "full rank" equal to the number of variables, but when one trait completely depends on another, the matrix has "deficient rank". Procrustes datasets always produce deficient rank covariance matrices due to translation, scaling, and rotation, but further deficiency will arise from trait integration. Furthermore, the eigenvalues of the matrix indicate whether covarying combinations of traits, each of which is associated with a major axis of variable space, are equally independent or whether some account for more of the total covariance than others, and thus whether the system is more modular or more tightly integrated. The 'full rank' of a matrix is the number of columns, while the real rank is the maximum number of linearly independent columns. Thus **K** is deficient because its real rank is less than its full rank. In this paper we utilize a third type of rank, the 'effective rank.' This is the number of significantly independent columns in the matrix, or the number of eigenvalues significantly different from zero. Here and in the following sections we discuss how these concepts have been used by previous authors to define indices of integration and modularity, and explain why they are insufficient.

Analysis of the phenotype is a multivariate problem (Cheverud, 1996). As stated by VanValen in the quote above, the shape space containing the objects of interest can be conceptualized as a hyperellipse in the original variable space. Use of principal components analysis ordinates this hyperellipse to the origin, and defines its principal axes. VanValen realized that this hyperellipse had two salient properties: 1) its dimensionality, and 2) its dispersion, which he called 'tightness'. In subsequent work the dispersion of the eigenvalues of a phenotypic covariance matrix—this tightness-- has received a great deal of attention, while its dimensionality has not (reviewed in Najarzadeh, 2019). Cheverud (1983) introduced the first modern measure of trait integration, defining it as

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

or the geometric mean of the eigenvalues of the correlation matrix subtracted from unity. This quantity is identical to one minus the 'standardized generalized variance', or SGV, in a wider statistical context, an elaboration of Wilks' generalized variance (1932) proposed by SenGupta (1987). The SGV is the geometric mean of the eigenvalue variance, and so may be thought of as the mean diameter of the axes in the hyperellipse. Both Cheverud (1983) and SenGupta (1987) state that the SGV is comparable among spaces of different dimensionality, and this assumption is widely accepted, although it has never been tested (e.g. Najarzadeh, 2019). Several other metrics have been proposed for quantifying the dispersion of phenotypic matrices, summarized by Pavlicev et al.,

2009, while the sample size requirements of various metrics are evaluated by Grabowski and Porto, 2017.

While both Cheverud's integration and the similar SVG have been used as measures of dispersion, a consensus has emerged that the standard deviation of the eigenvalues is a better measure statistically (Pavlicev et al., 2009), and this is the measure recommended by Goswami and Polly (2010). However, this eigenvalue standard deviation is routinely scaled to the mean eigenvalue, so that the original global variance magnitude is lost. This is advantageous when comparing matrices of different size, but causes information on the scale of variation to be lost. The SGV is not scaled in this way, but when Cheverud (1983) applied it to the correlation matrix there was a similar loss of relative variance magnitudes and inter-trait variance relationships as noted above. In this paper we use the eigenvalue standard deviation, and the SGV of the correlation matrix, as 'classical' integration measures to assess whole-shape integration. We demonstrate that both measures fail to accurately detect change in integration, and that they fail because they consider only variance dispersion, while ignoring the dimensionality dispersion, of the hyperellipse. The failure to consider dimensionality dispersion is exacerbated by the use of the correlation matrix.

Detecting and Testing Integration Patterns

"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.

The study of GPA covariance matrices expanded recently into the field of modularity and integration (Goswami et al. 2014; Klingenberg, 2013). This conception of integration is similar to Cheverud's in that it studies the covariance structure of **K**, however it differs by attempting to identify specific variable subsets that covary relative to other subsets. Rather than relying on a single overall measure of dispersion, modularity studies test specific models of landmark covariation against a null model in an attempt to characterize subsets with high intraset connectivity and low interset connectivity (Goswami and Polly, 2010). This approach has yielded powerful hypotheses of evolutionary plasticity and constraint linking development to phenotype, and to evolutionary changes in modularity over deep time (Goswami et al., 2015). Modularity models were first assessed using the RV coefficient introduced by Klingenberg (2008), a simple and flexible frequentist approach that tests for significant associations of proposed modules against a permuted null model. 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. Another common approach to assessing modularity model significance is Partial Least Squares (PLS; Goswami and Polly, 2010); this approach is more involved mathematically but has the same goal as the CR statistic (for a recent example on dog and wolf crania see Curth et al., 2017). Obviously, a major challenge in all modularity studies is initial specification of the model to be tested. There are several approaches to this, summarized by Goswami and Polly (2010), and those authors mention PCA as one useful method, although they do not explore it. In this paper

we use traditional morphometrics applied to interlandmark distances (Marcus, 1990) for initial exploration of covariance structure. We also use the multivariate generalization of the allometry equation (O'Keefe et al., 1999) to further examine the covariance structure. These analyses yield several module hypotheses that are then tested using the CR statistic.

As stated in the quote above, the modularity concept can also be applied to a whole shape. In whole-shape modularity, a shape would be more integrated if it had greater eigenvalue dispersion, and would be more modular if it possessed lesser dispersion. This is a specific prediction: in an evolving population, an increase in modularity should lead to a decrease in integration and hence dispersion, and thus a decrease in dispersion metrics like SVG and eigenvalue standard deviation. We test this prediction in an evolving population of dire wolves, and demonstrate that the eigenvalue standard deviation behaves in the opposite manner, while the SVG fails to detect the modularity change. The reason for the failure of these statistics is that they assume the number of significant principal components is the same in both populations, when it is not.

Information Entropy and Effective Rank

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.

The derivation of **K** directly from **X** (GPA coordinate data) results in a covariance matrix that is overdetermined, and this must be treated for in some way. Incidentally this problem is not unique to geometric morphometrics; the high degree of correlation in biological systems often results in matrices whose full rank is much higher than the real rank (Van Valen, 1974; Adams, 2016). The problem of rank deficiency can be framed in terms of the eigenvalues of a matrix, with the true rank represented by the subset of eigenvalues carrying significant variance, as opposed to the number that carry non-zero variance, which is how matrix rank is usually defined. The question of the number of 'significant' eigenvalues is a general question in data analysis, and there are many techniques for determining how many eigenvalues to retain. These methods are summarized by Cangelosi and Goriely (2007) in the context of cDNA microarray data, and 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. These are: 1) an integer value; 2) recourse to an *ad hoc* criterion to assess the cutoff for significance; and 3) space specificity, meaning the number of significant axes is specific to the space under study, and not comparable among spaces.

The number of 'significant' eigenvalues may be conceptualized in other ways; in terms of matrix algebra it is equivalent to some functional rank of **K** that is less than the full rank or the real rank, or to a measure of the rank deficiency of **K**. It is also equivalent to the number of dimensions present in **K**. The first to point out the problem of redundancy in the full and real ranks of **K** in a morphometric context was Van Valen

(1974), who tailored an information metric to correct for it. However the problem of redundancy in eigenvalue distributions is general, and has been solved in signal processing contexts by the use of information entropy (Shannon, 1948). Information entropy is a metric used to represent the information content of a system by determining how much a signal 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 was developed by Roy and Vetterli (2010); those authors term their metric the 'effective rank', and use it synonymously with 'effective dimensionality'. The following application of the Shannon entropy to geometric morphometrics follows the Roy and Vetterli (2010) development of effective rank.

Given a GPA covariance matrix **K** with eigenvalues $\mathbf{\Lambda_V}$:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_v \geq 0$$

The Shannon entropy E_S of **K** is defined as

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

Note that the eigenvalues are standardized to the trace of $\mathbf{\Lambda}$; 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; for the purposes of geometric morphometrics the 'effective rank' of **K** is defined in the same way:

$$R_e = \exp\,(E_S)$$

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 effective number of dimensions of the shape space represented by **K**. As we will show below, effective rank also suffers from the same problem as eigenvalue dispersion in that it is incapable of distinguishing differences in total variance. We will argue that it should be combined with the concept of SVG in a new metric that summarizes dispersion in both dimensionality and variance simultaneously.

Model System: Canis dirus *at Rancho La Brea*

The tar pits at Rancho La Brea (RLB) are an iconic example of a tar pit lagerstätten, and the type locality of the Rancholabrean Land Mammal Age. Over a century of excavation at RLB has yielded hundreds of thousands of fossils representing a diverse Late Pleistocene fauna. Carnivores are disproportionally represented at La Brea, forming 90% of historically recovered fossils, while 57% are from the genus *Canis* (and overwhelming from the dire wolf, *Canis dirus*; Stock and Harris, 1992). This wealth of RLB carnivore

fossils has attracted research attention for decades, and the deposit delineates our perception of the terminal Pleistocene in Southern California.

Background. — Carnivores have been the primary research subject at RLB given their ubiquity. The foundational study of Van Valkenburgh and Hertel (1993) on carnivore dentition was the first to establish that tooth breakage and tooth wear were elevated across the entire carnivore guild at RLB relative to those of modern communities. Further work also established that this elevated breakage and wear signal was heterogeneous, with dire wolves showing high, yet variable, breakage and wear, while that of *Smilodon* was less extreme and less variable (Binder and Van Valkenburgh, 2010). These studies focused attention on discrete breakage and wear events in the dire wolf, and led to attempts to capture evolutionary shape change associated with breakage patters. An emerging consensus on pit dating (Fuller et al., 2014) allowed investigation of evolutionary change in both dire wolves (O'Keefe et al., 2014; Brannick et al., 2015) and *Smilodon* (Meachen et al., 2014; Goswami et al., 2015). While these studies demonstrate conclusively that phenotypic evolution did occur in the dire wolf at RLB, correlations with possible drivers such as climate have been qualitative, as have analyses of the biotic processes underlying this evolution.

Previous work on dire wolves from Pits 13 and 61/67.— We compare the phenotypic covariance structures of two populations of dire wolves collected at RLB. The populations come from two pit deposits; one contains wolves enduring a severe breakage and wear event (Pit 13), while those from the other deposit are not (Pit 61/67; Figure 1; Binder et al., 2002). Elevated tooth breakage and wear tend to occur together and are

thought to indicate nutrient stress, as animals both process carcasses more completely, and compete more intensely, when prey resources are scarce (Van Valkenburgh, 1988, 2009; Meloro, 2012). Binder et al. (2002) first established that breakage and wear metrics vary significantly in dire wolves from different RLB pit deposits. Those authors also determined that this signal was not related to variable ontogenetic age, and therefore was ecologically driven. Subsequent work established this pattern beyond doubt, and was able to assign approximate ages to the relevant pit deposits (Binder and Van Valkenburgh, 2010; O'Keefe et al., 2014; Brannick et al., 2015). Surprisingly, elevated breakage and wear are absent in wolves just prior to the megafaunal mass extinction (Pit 61/67, deposited ~14 kya during the Bølling–Allerød interstadial). The deposit with the strongest breakage and wear signal is Pit 13, deposited circa 19 kya near the termination of the Last Glacial Maximum (O'Keefe et al., 2009; Fuller et al. 2014; O'Keefe et al., 2014).

Figure 1. Representative dire wolf dentaries from Pit 13 (A) and Pit 61/67 (B). Note relative crowding of teeth in the Pit 13 jaw, particularly in the premolar arcade, and wear on the molars and canine. The length of the carnassial is equal in the two jaws (35 mm) even though the bottom jaw is clearly larger overall. These morphologies lead O'Keefe et al. (2014) to conclude that Pit 13 wolves displayed ontogenetic stunting due to nutrient stress.

Morphometric analyses of dire wolf jaws (Brannick et al., 2015) and crania (O'Keefe et al., 2014) detected clear phenotypic differences in Pit 13 wolves relative to

those in 61/67. Pit 13 wolves are significantly smaller in overall body size, and also show significant shape differences, including diminution of the viscerocranium relative to the neurocranium. The decrease is size and associated allometric changes exhibited by Pit 13 wolves are consistent with ontogenetic stunting; they are the changes animals would display due to ontogenetic retardation due to chronic nutrient stress, as concluded by O'Keefe et al. (2014; Brannick et al., 2015). Assuming that elevated breakage and wear indicate nutrient stress in Pit 13 wolves, the observed size and shape changes were attributed to a chronic lack of nutrition during development, producing a neotenic adult phenotype. This interpretation is plausible; retardation of the positive ontogenetic allometry in the viscerocranium typical of all vertebrates (Goodrich, 1930) would produce the shape changes observed in Pit 13. Yet this interpretation has several limitations; first, analyses were qualitative, and did not attempt to quantify the stunting signal or formally establish its statistical significance. Second, and more troublesome, the attribution of the observed neotenic pattern to ontogenetic allometry only is suspect. Neotenic patterns are readily observable in both the static and evolutionary allometries of gray wolves (O'Keefe et al., 2013) and dire wolves (O'Keefe et al., 2014), so size diminution and associated neotenic shape change are not a unique signature of ontogenetic stunting. They may indicate increased mortality of larger wolves on short time scales (static allometry), or a decrease of mean body size through selection against large wolves (evolutionary allometry; for a review of allometric categories, see Klingenberg, 1996). Here we examine the assertion that the observed phenotype in Pit 13 is the result of ontogenetic allometry. Testing this assumption requires characterization of the covariance structure of Pit 13 wolves, and comparing it to wolves that are not

enduring a breakage and wear event—those of Pit 61/67, a descendant population deposited about five thousand years later.

Experimental Design

We compare the performance of eigenvalue standard deviation and SVG metrics in characterizing covariance structure using several different morphometric strategies applied to the dire wolf mandibles from Pit 13 and Pit 61/67. Methods used include distance-based principal components analysis to assess traditional multivariate allometry (Marcus, 1990; O'Keefe et al., 1999), landmark-based geometric morphometrics (Bookstein, 1997) and investigation of the resulting centroid size and relative warp variables, and evaluation of modularity and integration (Goswami and Polly, 2010), including two measures of whole-jaw integration: eigenvalue standard deviation and the SVG. We begin with an exploratory analysis of multivariate allometry vectors calculated from interlandmark distances. We proceed to univariate tests for shape differences between Pits 13 and 61/67. The multivariate allometry results also suggest several possible models of modularity, and these are tested using the CR statistic. Principal components analysis of the Procrustes-superimposed landmark data is then performed, allowing analyses of size and shape change between pits, and enumeration of the impact of static and ontogenetic allometry on these differences.

Evaluation of modularity models indicates that wolves from Pit 61/67 have a significantly modular covariance structure, while those from Pit 13 do not. Pit 13 wolves are therefore more integrated than those from 61/67, yet both classical measures of whole-shape modularity fail to capture this. Therefore we modify the SVG metric to

reflect the effective rank of the two populations. This new effective rank-standardized version of the SVG is successful in capturing the increase in integration in Pit 13 wolves. Variation in sample size is known to impact measures of integration (Grabowski and Porto, 2017), and we wish to control this as a nuisance parameter in this study. The samples sizes of the Pit 13 and Pit 61/67 allow this; the Pit 13 *n* is 36 wolves, while 61/67 contains 83. The large sample size in 61/67 allows us to jackknife it down to 36, furnishing a measure of dispersion for all statistics from 61/67 that Pit 13 can then be tested against. Sample size is therefore held constant initially in the experimental design. Lastly, we generalize the new metrics to the case of arbitrary sample size, and apply them to a data matrix from the genus *Smilodon* as a demonstration.

MATERIALS AND METHODS

Fossils utilized in this paper are left dentaries of *Canis dirus*, from Pits 13 and 61/67, housed in the Tar Pit Museum at Hancock Park, Los Angeles. The input data for all analyses were 14 landmarks taken from 119 dire wolf jaws, a subset of the 16 landmark data set from Brannick et al. (2015).

Data Collection

One hundred and nineteen adult *Canis dirus* dentaries were included in this study. The labial side of specimens from Pits 13 (n =36), and 61/67 (n = 83) were photographed using a tripod-mounted Cannon EOS 30D 8.20-megapixel camera. All specimens were anatomical lefts and were laid flat and photographed with a 5 cm scale-bar. While camera angle and distance were held constant when photographing, scale-bars were used to

properly size specimen images in Adobe Photoshop CS2 v.9.0 before landmark digitization; scaling was mostly unnecessary, and always minor.

Fourteen homologous landmarks were digitized on each specimen using the program tpsDig2 (Rohlf, 2013). Positions of landmarks were chosen to give a general outline of the mandible and capture information of functional relevance (Brannick et al., 2015; Figure 2A). Landmarks on the tooth row were placed on the alveolus so specimens with missing teeth could be included. However, presence of the lower carnassial tooth was required in order to obtain landmark 5. Landmark 5 distinguishes the trigonid basin from the talonid basin. Two landmarks from Brannick et al. were not used; one of these, at the angle of the mandible, was difficult to identify and showed unacceptably large variance relative to the other landmarks. The second was an interior landmark in the masseteric fossa; it was deleted to make the dataset amenable to outline-based iterative semilandmarking.

Multivariate Allometry

Initial analysis consisted of calculating the interlandmark distances from the raw digitized landmark data **LM** (Figure 1B). These distances were then evaluated for multivariate allometry by calculating the first principal component vector of the covariance matrix of *ln*-transformed distances; this procedure yields the multivariate allometry vector (Jolicoeur, 1963; O'Keefe et al., 1999; O'Keefe et al., 2013). Confidence intervals for the allometry vector coefficients where calculated by bootstrapping the input data (1000 replicates) in the R core package (R Core Team, 2014). The eigenvalues of the full 13-

measure multivariate allometry vector were unusual, necessitating additional runs with some variables removed; all runs are reported in Table 1.

ANOVAs, Procrustes Superimposition, and Principal Components

In order to evaluate differences in the sizes of individual teeth between Pits 13 and 61/67, two-tailed analyses of variance were run on the anterior-posterior width of the canine (landmark 1 to 2), the premolar arcade (landmark 3 to 4), and the carnassial (landmarks 4 to 6). Because Pit 13 wolves are smaller than those from 61/67 (O'Keefe et al. 2014, Brannick et al. 2015), the distances were divided by the geometric mean of each row as an isometric size correction (ANOVAs on raw distances are significantly larger for Pit 61/67 wolves in all cases). ANOVAs of these distances, and of the geometric mean itself, can be found in Table 2. All ANOVAs were executed in the statistical program JMP.

Standard Procrustes superimposition of the landmark data was performed using the **geomorph** package (Adams et al., 2020) in R (R Core Team 2014). Procrustes superimposition yielded a matrix of transformed (GPA) landmarks **X** and a vector of centroid size. Principal components of three matrices were calculated from these data: $\mathbf{X_{13}}$, $\mathbf{X_{6167}}$, and $\mathbf{X_{all}}$. The initial pooled analysis used the correlation matrix and was used for visualization and relative warp analysis, while classical integration metrics were run on correlation matrices of each pit sample. All information entropy-based metrics were run on the covariance matrices. ANOVAs on all pooled principal components were performed in JMP and are reported in Table 2. A plot of the pooled principal component scores can be found in Figure 3, while lattice diagrams illustrating relative warping on PC I and PC II are depicted in Figure 4.

Modularity Model Tests

Six models of modularity were suggested from the multivariate allometry analyses of the interlandmark distances, and these models were tested individually using the CR statistic (Goswami and Polly, 2010), 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; all tests and results are listed in Table 3.

Measures of Whole-Jaw Integration

Classical measures and effective rank. — Whole-jaw integration was measured in three ways. The first two are classical dispersion measures: eigenvalue dispersion and SGV. Eigenvalue dispersion was measured as the standard deviation of the eigenvalue vector of the correlation matrix of GPA landmarks from each pit, standardized to the mean eigenvector (Goswami and Polly, 2010, modified from Equations 7 and 8):

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

We note that we considered only the first 24 eigenvalues, as four degrees of freedom are lost to the Procrustes analysis. Eigenvalue dispersion was calculated using eigenanalysis, bootstrapping, and permutation codes written in R (R Core Team, 2014).

Confidence intervals for Pit 61/67 were derived from 10000 jackknife replicates, each downsampled from 83 to 36 individuals. The second measure of dispersion was the SGV, calculated as the 24th root of the first 24 eigenvectors of the correlation matrix to account for the loss of four degrees of freedom to the GPA:

$$SGV_{24} = \sqrt[24]{\prod_{i=1}^{24} \lambda_i}$$

with confidence intervals for Pit 61/67 jackknifed in the same way as the eigenvalue standard deviation. Values for eigenvalue standard deviation and for SGV_{24} can be found in Table 4. Table 4 also includes values for the effective rank of each data partition, calculated using the formula in the Introduction and using code written in R. Effective ranks are illustrated in Figure 5, along with standardized and non-standardized scree plots of the eigenvalues of each data partition. Confidence intervals for Pit 61/67 were calculated from 10000 jackknife downsamples to an *n* of 36. Lastly, Table 4 also includes the trace of $\mathbf{\Lambda}$ for each of the data partitions, accompanied by jackknifed confidence intervals for Pit 61/67.

Effective rank-scaled SGV or 'effective dispersion'.— To account for deficiencies in the other to metrics, we introduce a third measure of whole jaw integration, which is the SGV scaled by the effective rank. We label this quantity SGV_{Re}, and calculate it as follows:

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

Where R_e is the effective rank of the covariance matrix of the data partition, and $[R_e]$ is the integer value of the effective rank. This coefficient can be described at 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. This metric accounts for both the dispersion of eigenvalue variance, and the dispersion of eigenvalue dimensionality, in a single number. 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.

Previous authors rejected SGV as a statistic to measure dispersion because its distribution is not linear (Pavlicev et al., 2008); they prefered 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 a measure of variance, one may utilize the definition of the standard deviation to transform the SGV_{Re} into a similar measure:

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

where the square root of the SGV_{Re} is a new metric, D_e, that we call ‘effective dispersion’ because it measures dispersion in both variance and dimensionality together, which are jointly necessary to describe morphological integration and modularity.

Matrix Size and Relative Dispersion

All whole-shape integration measures calculated here have ignored the impact of sample size. Sample size clearly affects 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 (Table 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. (2009) 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 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. One can generate a quantity with these properties by permuting the columns in **X** and then calculating the resulting covariance matrix. This matrix will preserve the input variances on the diagonal, but will remove statistical correlation on the off-diagonal (although covariances will still be non-zero, Pavlicev et al., 2009). 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, and can be used as a basis for standardizing D_e for matrix size.

The appropriate scaling is based on consideration of what the effective dispersion actually is, namely the square root of the geometric mean of a subset of the eigenvalues of **K**, or alternatively, the root of the variance of one dimension of the hyperellipse. Because D_e it is the square root of a mean eigenvalue (a variance), it has the same units as the original data (Pavlicev et al, 2009), and we would like to conserve this property in the scaled coefficient. Therefore there should be no variance term in the standardization on dimensional grounds. More importantly, the other property of D_e is that it has been limited to its non-redundant information content, and we wish to scale D_e to an expectation of maximum possible information content. 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:

$$D_r = \frac{D_e}{\sqrt{\frac{1}{\pi R_e}}}$$

The column values in each landmark matrix were permuted 10000 times using code written in R. The permuted effective ranks are reported in Table 4, as is the relative dispersion. The relative dispersion is comparable among data sets of different sample

sizes, as demonstrated in Table 4 (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 0.0238).

Lastly, as a demonstration of the utility of the relative dispersion metric, we analyze a data set of *Smilodon fatalis* (Meachen et al., 2014). The data set comprised 14 landmarks digitized from photographs of 81 *Smilodon* jaws originating from pits 61/67, 13, and the UCMP collection. Pooling of the *Smilodon* sample was necessary to increase sample size; we note that time averaging may be a confounding factor here. All relevant statistics for this *Smilodon* are also listed in Table 4, while a scree plot with bootstrapped effective and permuted ranks of both Smilodon and Pit 61/67 dire wolves is presented in Figure 6.

Results

Multivariate allometry and univariate ANOVAs

The multivariate allometry vector for all 13 interlandmark distances is odd, with a single variable scoring very highly (the distance between the canine and the first premolar, landmarks 3-4), while the rest show significantly negative coefficients (Table 1, Figure 1). This pattern indicates that the c1-p1 distance is highly variable, and that this variance is not correlated with any other variable. Further analysis shows that the c1-p1 distance crosses the boundary between two modules, and its variance arises from at least two allometric factors (discussed below), accounting for its unique loading. We therefore discarded that distance and ran the analysis again with 12 distances. The resulting vector is strongly positively allometric for two variables, the anteroposterior width of the canine (1-2) and the length of the condyloid process (9-10). Lengths associated with the other

teeth are negatively allometric (Figure 1B). The covariation of canine width and condyloid process length suggests a functional unit, and provides one hypothesis for a testable module in the modularity tests. The lengths of the other teeth are all negatively allometric, as are the landmarks on the inferior margin on the jaw that are linked to tooth location. This finding of negative dental allometry in the cheek teeth has been documented before, both in *Canis lupus* (O'Keefe et al., 2013) and in *Canis dirus* (O'Keefe et al., 2014). Negative dental allometry is a consequence of the adult teeth forming early in ontogeny, erupting before the end of the first year, while full somatic growth is not attained until the end of the second year (O'Keefe et al., 2013).

Interlandmark Distance	**13**	**+/-**	**12**	**+/-**	**10**	**+/-**
1-2	0.0656	0.0412	0.4889	0.0561	-	-
2-3	0.9455	0.0208	-	-	-	-
3-4	0.0259	0.0167	0.1883	0.0174	0.2152	0.0399
4-6	0.0132	0.0101	0.1500	0.0262	0.1783	0.0487
6-7	0.0175	0.0135	0.1439	0.0597	0.2341	0.1051
7-8	0.0704	0.0182	0.1927	0.0201	0.2895	0.0327
8-9	0.0720	0.0183	0.2043	0.0179	0.3119	0.0295
9-10	0.1188	0.0530	0.5682	0.0425	-	-
10-11	0.1073	0.0270	0.2711	0.0327	0.4071	0.0632
11-12	0.0759	0.0193	0.2105	0.0226	0.3004	0.0321

12-13	0.0134	0.0105	0.1531	0.0285	0.1790	0.0525
13-14	0.0250	0.0163	0.1842	0.0196	0.2098	0.0419
14-1	0.2215	0.0389	0.2967	0.0659	0.5623	0.1065
Isometry	0.277		0.289		0.316	

Table 1. Allometry vectors calculated from interlandmark distances on *Canis dirus* dentaries as shown in Figure 1. Three vectors are shown; the first includes all interlandmark distances, the second omits the distance between c1 and p1, and the third omits this distance as well as the width of c1 and the length of the condyloid process. The isometric coefficient value for each vector is shown at bottom; confidence intervals for coefficient values are one standard deviation of 1000 bootstrap replicates of each vector. Positively allometric values are shown in red, negative in blue, and isometric in black.

To further illustrate this pattern, we ran a third allometry vector that excluded both the c1-p1 distance (2-3), and the two positively allometric measures (1-2, 9-10). This vector clearly shows the negative allometry of the premolars and molars, and either isometry or positive allometry in distances that span the jaw corpus (Figure 1D). This negative dental allometry in the cheek teeth suggests a second module, comprising the lengths of the premolar arcade, the carnassial, and m2.

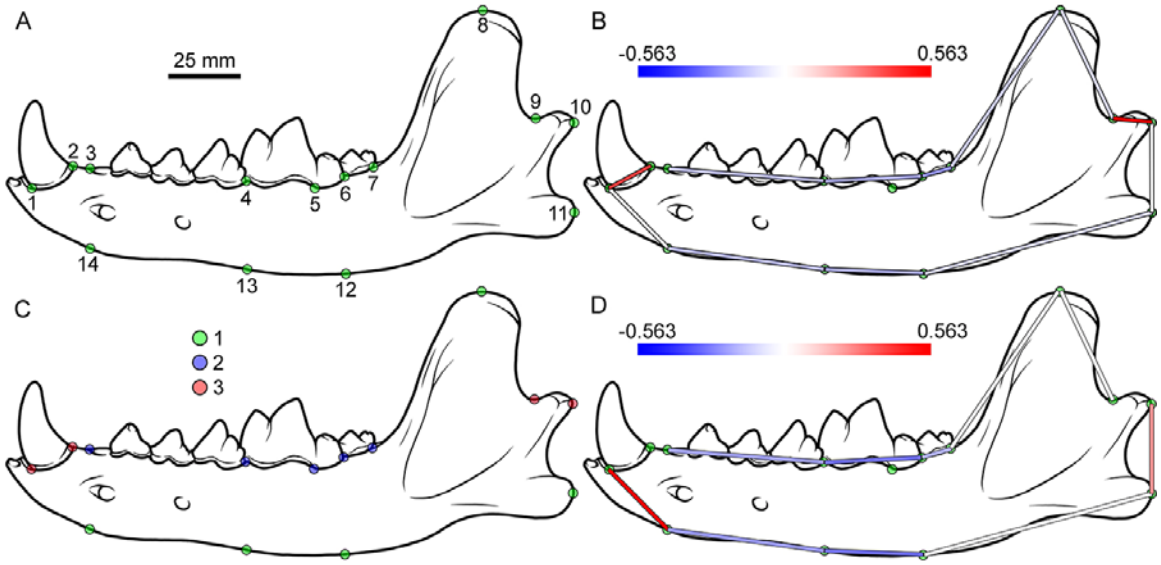

Figure 2. Locations of the 14 landmarks used in this study, placed on an outline drawing made from a digitized photograph of an RLB *Canis dirus* mandible (A). Interlandmark distances shown in B and D are those used in multivariate allometry and univariate tests. Panel B depicts the 12-measure multivariate allometry vector, with each length heat mapped to its allometry coefficient from Table 1. Panel D shows the same for the 10-measure allometry vector. Landmark memberships in C are modular groups identified in Pit 61/67 wolves, tested and summarized in Table 3.

If the Pit 13 wolves experienced ontogenetic stunting due to nutrient stress, their teeth should be more crowded in the jaw due to retardation of the growth of the jaw corpus. This is easily observable by eye (Figure 1); The teeth are essentially the same size in the two wolves, while the jaw corpus is shorter in the Pit 13 wolf, and the premolar arcade is more crowded. Lastly, the Pit 13 wolf shows significant tooth wear, including on the canine, while the other does not. In summary, multivariate allometry calculations readily identify the pattern of late-stage ontogenetic growth (positive allometry in the jaw

corpus versus the dentition) in dire wolves previously identified by O'Keefe et al. 2014, and O'Keefe and Brannick, 2015. The multivariate allometry calculations also demonstrate that the canine does not covary with the cheek teeth; instead it is strongly positively allometric, and correlated with strong positive allometry in the condyloid process. This suggests both a biomechanical linkage, and a module worth testing. Lastly, the teeth should not be treated as a single module in canids (e.g. Segura et al., 2020).

Measures of overall size consistently find that Pit 13 wolves average smaller than those from 61/67, a signal first recovered here by an ANOVA on the geometric mean of the interlandmark distances (Table 2). Pit 61/67 wolves average larger at a high level of significance ($p < 0.0001$). This finding means that at least some of the size variation between pits is not a result of late-growth neoteny. If that were the case, the teeth would not differ in size between pits. To investigate patterns of relative tooth sizes between pits we corrected for size by dividing interlandmark distances by the geometric mean. These data demonstrate that the canine is relatively much larger in 61/67 wolves, while the carnassial is relatively larger in Pit 13 wolves. These decoupled allometries between the carnassial and the canine are also clear in the multivariate allometry vectors, and are not an obvious result of ontogenetic stunting. Both imply the teeth comprise more than one module, and should not be tested as one.

Variable	**ANOVA**
Interlandmark Geometric Mean	<0.0001* 61/67 > 13
Canine size = 1-2/GM	<0.0001* 61/67 > 13
Premolar Arcade = 3-4/GM	0.3228

Carnassial = 4-6/GM	0.0021* 13 > 61/67
m2 = 6-7/GM	0.1241
Centroid Size	<0.0001* 61/67 > 13
PC I (20.2 %)	0.4970
PC II (15.3 %)	0.0059* 61/67 > 13
PC III (12.2 %)	0.0518~ 61/67 > 13

Table 2. Summary of t-test on means of two-way ANOVA with pit of origin as the independent variable, for the variables shown. Percent of variance explained is listed after each PCA vector in parentheses; the fourth and higher components accounted for less than 8% of the variance in both analyses, and are not considered further.

Procrustes Superimposition and Principal Components

Initial Procrustes superimposition creates a vector of centroid size, subjected to an ANOVA by pit in Table 2. Again Pit 61/67 wolves are found to be larger than those from Pit 13 at a high level of significance ($p < 0.0001$). Eigenanalysis of the correlation matrix of GPA landmarks yields an ordination whose first six eigenvalues explain 20.2%, 15.3%, 12.2%, 7.4%, 7.3%, and 5.6% of the total variance. We restrict our examination to the first three principal components based on the large variance drop between eigenvalues three and four. Principal components I, II, and III were subjected to ANOVAs by pit (Table 2); PC I does not segregate the pits at all, while the pits are significantly different on PC II ($p < 0.0059$) and marginally different on PC III ($p < 0.0518$). Interestingly, PC I is correlated with centroid size ($r^2 = 0.142$) even though it does not discriminate between pits. It therefore carries an allometric signal that is

common to the two deposits. Principal component II is also weakly correlated with size ($r^2 = 0.073$) yet discriminates strongly between pits, so should give an indication of how the two samples differ.

We also explored *a priori* correction for allometry before ordination by rerunning the analyses using the residuals of regression against centroid size (Drake and Klingenberg, 2010). The results of this ordination were similar to the one described above; the first six eigenvalues were 19.2%, 14.5%, 12.7%, 7.7%, 7.2%, and 5.5%. Analysis of variance on PC I-III show that both PC I and III are marginally significant discriminators while PC II is not. This ordination is harder to interpret, with inter-pit variance spread weakly across two axes. We therefore concentrate on interpretation of the first ordination. This ordination contains the size factor, and that signal largely drives PC I, allowing its *a posteriori* identification.

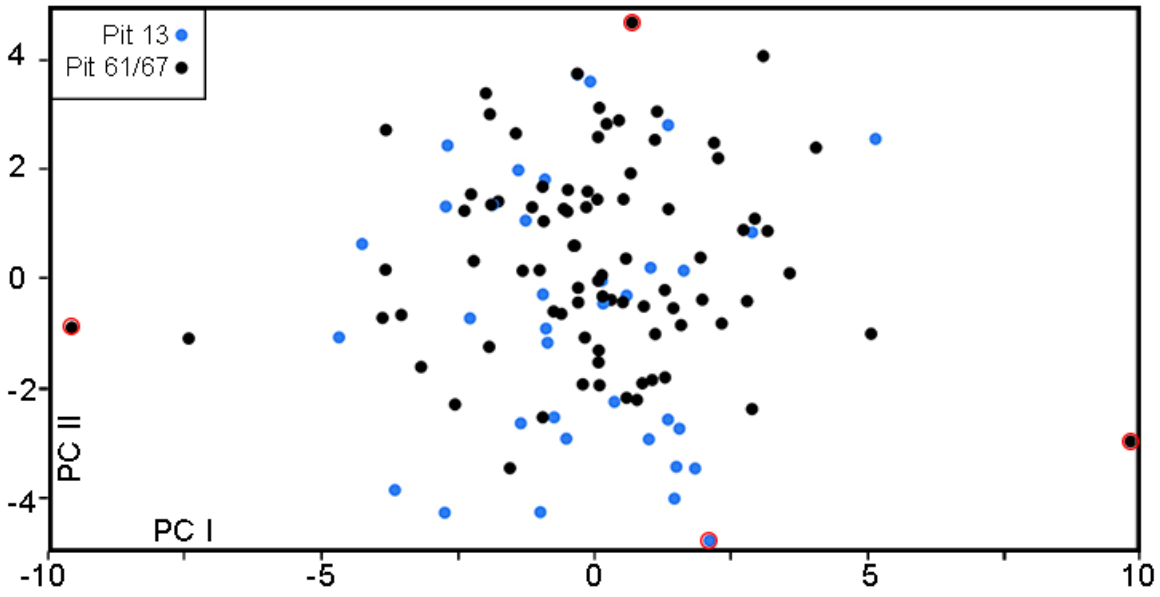

Figure 3. Scatter plot of PCI vs. PC II of the pooled PCA for La Brea dire wolves. Axes are scaled to percent variance explained. Pit 13 wolves are in blue. These

eigenvectors are from the raw Procrustes-superimposed landmarks, without treatment of allometry. The retention of size-correlated signal aids in interpretation of the PC axes; for further discussion see text. Points circled in red indicate jaws depicted in Figure 4 to illustrate variation along each vector.

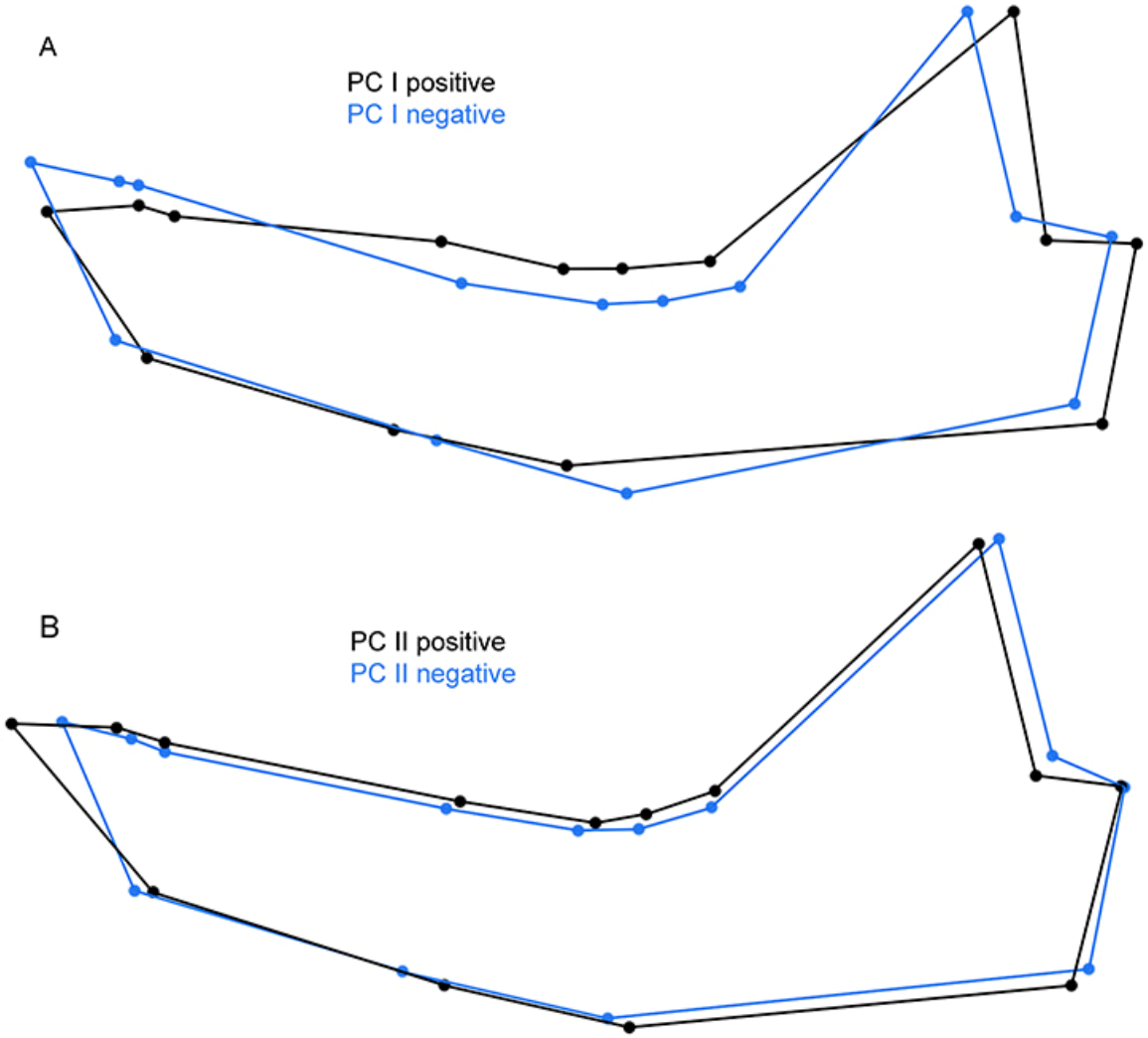

Figure 4. Overlays of landmark positions for extreme specimens on PC I (A) and PC II (B). Principal component I is highly correlated with centroid size, but does not differentiate between pits. It carries a signal of jaw corpus size increase relative to the cheek teeth. Principal component 2 is also correlated with centroid size and is a highly significant discriminator between Pits 13 and 61/67. It carries a signal of an increase in

canine width and condyloid process length. Lattices are plotted from the specimens circled in red in Figure 3.

Interpretation of PC I is aided by examination of the lattice diagrams in Figure 4a. Jaws scoring positively on this axis are larger overall, due to the correlation with centroid size. In terms of shape, the corpus of the jaw is dorso-ventrally wide in large animals, while relative expansion of the coronoid, condyloid, and angular processes indicate the osseous part of the jaw is larger on positively-scoring jaws as well. The premolar arcade is relatively short in jaws scoring positive on this axis, while the canine is rotated to a more procumbent position. The canine is not significantly larger on PC1 even though that axis is highly correlated with size. All of these shape differences are readily interpretable as effects of the bony jaw growing up around the teeth during late stage growth, a factor that is also readily apparent in analyses of *Canis lupus* crania and jaws (O'Keefe et al. 2013). Importantly, Because PCI does not discriminate between Pit 13 and Pit 61/67, this result falsifies the hypothesis that the two populations differ solely through the presence of neotenic stunting of ontogenetic allometry (contra O'Keefe et al., 2014). This effect is a prominent feature of the covariance structure but it does not account for the shape difference between pits; it is similar in both populations, and does not discriminate between them.

In contrast, Principal Component II is a good discriminator between pits. These jaws are similar except for two areas: the size of the canine and the length of the condyloid process. This correlation was also identified in the allometry vectors reported above, and it is uncorrelated with the ontogenetic stunting signal carried on PC I. The

appearance of this signal on PC II indicates that this factor is the main discriminator between the two wolf samples and highlights the need to analyze it further as a module within the jaw, as the canine and the condyloid process appear to covary. This unit displays marked positive allometry, increasing greatly with body size. Anterior rostrum width is driven by canine size, and this also shows strong positive static allometry in gray wolves (O'Keefe et al., 2013). The primary shape difference between pits is therefore an allometric diminution of the canine linked to small body size. This allometry may be static, evolutionary, or both.

Modularity Model Tests

A range of modularity models were run on the GPA data in an attempt to further characterize patterns of covariation. These tests were initially run on the pooled data set, then on the data split between pits. The salient finding here is that Pit 13 wolves display no significant modularity, while the 61/67 wolves are modular (Table 3). 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. Given that 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 running the CR analysis. Although confidence intervals widened in these subsets, the covariance structure of Pit 61/67 was still clearly evident and significantly modular, indicating that the modularity difference between pits is not attributable to difference in sample size. The pooled CR statistics, run on all 119 wolves, indicate that the inclusion of

the Pit 13 wolves actually degrades the global modularity signal. This demonstrates that the relative lack of modularity in Pit 13 is real.

Model	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; 3,4,5,6,7	1.06	0.244	1.192	0.719	1.029	0.226	-
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	1.10	0.573	1.201	0.838	1.090	0.491	-

Table 3. Results of modularity model tests. Module models are listed on the left column, with CR statistics for the three data partitions in subsequent columns. Pit 61/67 wolves have one significant module comprising the check teeth, while Pit 13 wolves do not.

Models with statistically significant CR values include two module models that contrast the length of the post-canine tooth row with the rest of the jaw, and a three-module model comprising the cheek tooth row, the width of the canine, and the rest of the jaw. The effect sizes of the three 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. For Pit 61/67 wolves only, the cheek tooth model was significantly better than

the null model, while the model containing only the molars was marginally so. The anterior-posterior width of the canine and of the condyloid process 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 cheek teeth are clearly a module, and the positive allometry of the bony portion of the jaw is evident (although not significantly in Pit 13). 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. 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 integrated than Pit 61/67 wolves.

Measures of Whole-Jaw Integration

The eigenvalue standard deviation metric is significantly less in Pit 13 wolves (Table 4: $\lambda\sigma$; 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 all 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 classically-measured eigenvalue dispersion (measured as eigenvalue standard deviation) 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 below, over half of the ranks in that metric are meaningless, and we believe the SGV_{24} is essentially measuring

noise. We experimented with the use of the covariance matrix in the SGV_{24}, but this also failed, as the metric was driven by the higher total input variance in the Pit 13 matrix. This covariance-based metric may be useable if standardized to the input variance magnitude, but will still require a sample size correction, and will still suffer from the inclusion of many meaningless, near-zero eigenvalues.

Calculation of the effective rank for the two populations indicates that the dimensionality of Pit 61/67 is less than that of Pit 13 (Table 4: R_e, 13 > 61/67, p = 0.00148). This difference in dimensionality dispersion is highly significant. We therefore employed the R_e scaled SGV developed above, and tested the pits again (Table 4: SGV_{Re}, 13 > 61/67, p = 0.0137). Using this measure, Pit 13 wolves have 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.

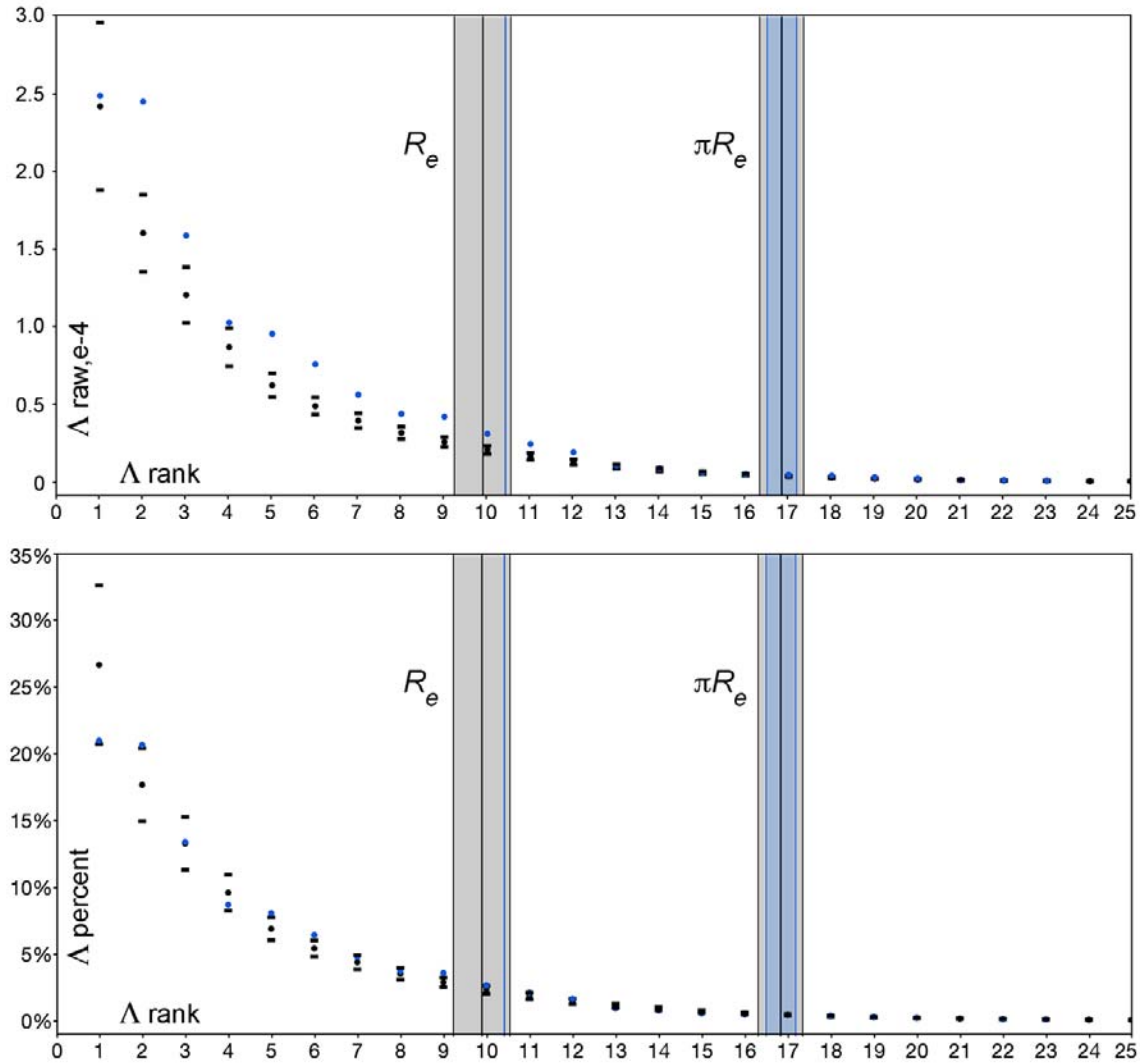

Figure 5. Scree plots and effective rank for GPA landmark data of *Canis dirus* jaws. Blue is Pit 13, black is Pit 61/67. The top pane depicts raw eigenvalues, the bottom eigenvalues standardized to percent variance explained. Confidence intervals for effective rank (R_e) are 10000 jackknife replicates of the Pit 61/67 sample. Eigenvalue dispersion is obviously greater in Pit 61/67 wolves in the top graph, 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 larger, so the variance is spread over a greater number of dimensions. Calculations of effective rank requires standardization of the sum

of the eigenvalues to unity, and is the same in both plots. The permuted dimensionality is also shown on both plots; this value is identical between pits (πR_e = 16.84), demonstrating that the dimensionality difference between the pits is a real biological signal. The permuted dimensionality also demonstrates that the maximum full rank of the input data is overdetermined, while the real rank of the data is highly overdetermined, with fewer than half of the 24 eigenvalues being meaningful.

Another scree plot for *Canis dirus*, this time jackknifed to an *n* = 81, and also including a scree plot for *Smilodon*, is shown in Figure 6. We standardized the sample size to 81 for both taxa for this comparison and for the bootstrapped confidence intervals reported in the caption for Figure 6. *Smilodon* and *Canis dirus* are similar in that both matrices are highly overdetermined, 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 a sample size of 36. This implies that the permuted rank of the matrices should converge toward 24 at large *n* (probably over 100, in accord with sample size requirements of other integration metrics as shown by Grabowski and Porto, 2017).

Data Partition	$\lambda\sigma$, +/-	$\Sigma\lambda$, +/-	R_e, +/-	πR_e, +/-	SGV_{24}, +/-	SGV_{Re}, +/-	D_e	D_r
C. dirus 13	0.318	1.19e-3	10.414	16.841	0.362	4.131e-5	0.0064	0.0263
n = 36				0.345				
C. dirus 6167j	0.348	9.06e-4	9.89	16.832	0.361	3.490e-5	0.0058	0.0238
n = 36	0.030	0.911	0.656	0.519	0.029	1.42	0.0012	0.0049

C. dirus 6167	0.319	9.06e-4	11.362	21.028	0.500	2.71e-5	0.0052	0.0238
n = 83				0.226				
Smilodon	0.287	2.49e-3	12.32	21.122	0.575	7.20e-5	0.0085	0.0400
n = 81				0.228				

Table 4. Whole jaw integration measures and other statistics for the *Canis dirus* mandible data partitions. All reported confidence intervals are one standard deviation of 10000 jackknife or permutation replicates. Quantities listed are eigenvalue standard deviation (correlation matrix, 24 eigenvalues), $(\lambda\sigma)$; sum of eigenvalue variance, $\Sigma\lambda$; effective rank, (R_e); permuted effective rank, (πR_e); standardized generalized variance 24 (SGV_{24}); standardized generalized variance effective rank (SGV_{Re}); effective dispersion (D_e); and relative dispersion (D_r).

Centroid Size

Table 5 lists the pertinent statistics for centroid size. All moments for centroid size in this table were calculated from the full *n* of each data partition. As state above the mean for 61/67 is greater; however the range and standard deviation for that pit are also significantly higher than for Pit 13. Pits 13 is heavily right skewed, and its kurtosis is much higher, meaning the 13 distribution skews small and varies less that the 61/67 distribution. The decrease in size and shape variance in Pit 13 implies selection on body size (Lande, 1980; Melo and Marroig, 2015).

Data Set	$CS\mu$	$CS\sigma$	*CSskew*	*CSkurt*	*CScv*
C. dirus 13	2542.3	91.25	0.713	0.304	3.59

n = 36					
C. dirus 61/67 *n* = 83	2704.6	121.63	-0.208	-0.609	4.50
Smilodon *n* = 81	2826.7	170.72	-0.50	0.148	6.04

Table 5. Descriptive statistics for the centroid size distributions. Moments listed are the mean(***CSμ***), standard deviation(***CSσ***), skewness(***CSskew***), kurtosis(***CSkurt***), and coefficient of variation (***CScv***) of the centroid size vector recovered from the global PCA for *Canis dirus*; *Smilodon* was a separate PCA.

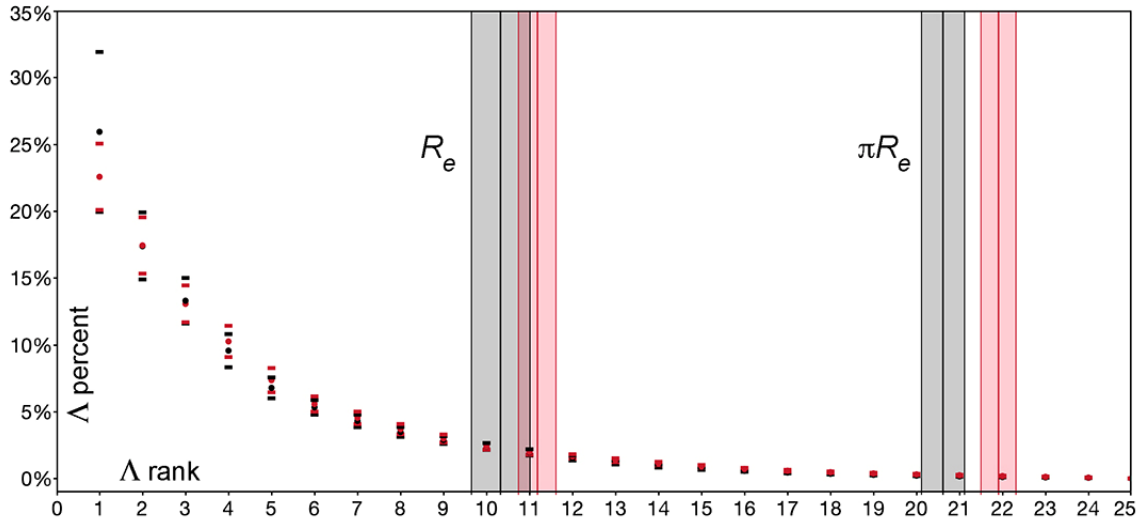

Figure 6. Comparison scree plots for *Canis dirus* and *Smilodon* data sets, 14 landmarks on jaws, at *n* = 81. Plotting conventions are the same as those used in Figure 5. *Canis dirus* is shown in black, *Smilodon* in red. Both data sets have been bootstrapped for a confidence interval (10000 replicates), hence the rank values shown here are less than the precise values shown in Table 4. This illustrates the strong sensitivity of the effective rank calculation to sample size. The plotted means and standard deviations are: *Canis*

dirus 61/67, $\boldsymbol{R_e}$ = 10.33 +/- 0.677; $\boldsymbol{\pi R_e}$ = 20.614 +/-0.503. Smilodon $\boldsymbol{R_e}$ = 11.184 +/- 0.438; $\boldsymbol{\pi R_e}$ = 21.91 +/- 0.503.

Discussion

We break the discussion into two main topics. The first is the biological interpretation of the evolutionary change between Pit 13 dire wolves and their descendants in Pit 61/67. The major shape change factor is probably static allometry linked to body size change. The second topic is the failure of classical whole shape integration metrics to capture the evolutionary increase in modularity, and the theoretical implications of the effective rank scaling of eigenvalue dispersion.

Evolutionary Change in RLB Dire Wolves

The phenotypic change in Pit 61/67 wolves relative to their ancestors in Pit 13 involves both size and shape, and hence must include allometry (Klingenberg, 2013). Pit 61/67 wolves are larger than those in 13. This is a highly significant finding recovered in ANOVAs of both the geometric mean of interlandmark distances and of GPA-derived centroid size. However, the hypothesis that this size difference is due to neoteny via stunting of late-stage growth is not supported. The basic structure of phenotypic covariance in dire wolf jaws closely resembles that of the gray wolf. It is highly allometric, expressing a strong signal of late-stage somatic growth due to early eruption of the adult dentition (O'Keefe et al., 2013). This ontogenetic allometry signal is one of two primary findings of the multivariate allometry analysis. It is also captured on PC I of the ordination of PGA landmarks; shape change along this component clearly shows the

bony jaw growing up around the teeth (Figure 4A). This component does not discriminate between pits, even though it is correlated with size. Instead this allometric factor is common to Pit 13 and 61/67 wolves. In retrospect this is not surprising; if the shape difference between pits were due only to retardation along this factor, the individual teeth would not vary in size between pits. All teeth are in fact smaller in Pit 13 wolves, meaning that if there is a signal of retardation along this axis, its footprint is too subtle to explain the gross size difference between pits, and that ontogenetic allometry is not the primary factor driving evolutionary change.

The second factor present in the *Canis dirus* covariance structure is strong positive allometry in canine size, coupled to an increase in the length of the condyloid process. This signal is clear in the multivariate allometry vectors, and larger wolves have relatively much larger canines, a strong pattern that is shown clearly in the univariate interlandmark distance tests. This signal also separates cleanly on PC II; wolves with relatively large canines and long condyloid processes score positively on this axis. Additionally, PC II is a strong discriminator between pits, a logical pattern given that Pit 61/67 wolves exhibit larger body size. Pit 13 wolves are smaller, and this is an evolutionary change in at least the middle term; it cannot be explained solely by growth retardation, but requires an evolutionary change toward smaller body size. This change could result from pure static allometry, where only smaller wolves survive during Pit 13 deposition; or it could result from evolutionary allometry, where selection for smaller body size is driving the population down the static allometry axis. These possibilities are not mutually exclusive so both may be operating. Yet there is no wholesale reorganization of the dire wolf covariance structure between Pits 13 and 61/67. The

positive allometry in canine size is a constant, in dire wolves and in canids generally. Relatively small canines are a consequence of smaller body size, which is a real factor given the diminutive size of all teeth in Pit 13 wolves. Based on examination of the centroid size vector, particularly the narrowing of its dispersion and its pronounced right skew, we hypothesize that Pit 13 wolves were under active selection for small body size. The population is pushed down the static allometry axis, and the canines decrease in size. Yet to incur heavy wear, an animal must survive for a long time. An evolutionary scenario of diet switching is plausible for RLB *Canis dirus*: perhaps Pit 13 wolves ate more bones than those consumed by 61/67 wolves, and a smaller body size was advantageous to acquiring this tough resource. These questions and hypotheses are currently under investigation using stable isotopes and microwear analyses, by the authors and others.

The global PCA results further indicate that wolves from both pits display the ontogenetic allometry signal. Why then did the modularity model tests not identify the cheek teeth module in Pit 13? The *p*-value for this module was 0.11, so those data have some tendency to show this module, but it does not rise to significance. We believe this is another impact of small body size; all allometries will be more weakly expressed across a smaller size range, particularly one that is narrower and skewed toward small animals. The Pit 13 ellipse is of smaller diameter, probably driven by selection for small body size. The associated allometric shape and modularity changes arise from the population moving down the static allometry axis.

Whole Shape Integration Measures

This paper demonstrates that classical measures of whole-shape integration do not perform as expected when applied to the *Canis dirus* test case. They fail for two reasons. The first is that they rely on the correlation matrix, and this matrix inflates trivial variance from small coordinate variables. It also inflates trivial variance from meaningless components, because geometric morphometric matrices are overdetermined. The second reason is that the two metrics measure variance dispersion, but ignore dimensionality dispersion, and so are misled by data sets that differ in dimensionality. Using the concept of information entropy, we show that the dimensionality of Pit 13 wolves is significantly greater. Calculation of the covariance matrix-based SGV metric using the effective rank, rather than the full rank, is suggested to accommodate both types of dispersion. We demonstrate that the effective rank scaled SGV metric successfully recovers the evolution of increased modularity in Pit 61/67 wolves. The behavior of this metric must be examined, and a simulation study using matrices of known structure is a logical next step. Examination of Table 4 shows that effective rank is very sensitive to sample size, as are the other metrics. Therefore we introduce the concept of relative dispersion, D_r, to account for matrix size. Relative dispersion relies on the covariance matrix, so the input variance is not distorted as are metrics relying on the correlation matrix. It accounts for dispersion in dimensionality, as well as dispersion in variance. And it accounts for differences in matrix size by standardizing against the permuted dimensionality. 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

As demonstrated in Figure 6, The relative dispersion has application in the study of whole-object dimensionality of biological shape as represented by landmark data. The relative dispersion has these advantages: 1) it is continuous rather than discrete, and 2) does not rely on *ad hoc* criteria for a dimensionality cutoff. However, this metric still suffers from the third limitation on hyperellipsoid characterization, that of space specificity. 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 (11 of the 14 landmarks are in fact homologous). Yet this is not true. A landmark-defined space with covariance matrix $\mathbf{K_V}$ 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 is negligible, nor that the proportion $\mathbf{K_R}/\mathbf{K_T}$ 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

$$K_T \approx K_V, and\ K_R \approx 0$$

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. Using this procedure allows definition 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 effective dispersion of a matrix that is semilandmarked densely enough to asymptotically stabilize D_r. Classical integration measures are not amenable to this procedure, because they consider all eigenvalues, and as the number of landmarks increases the rank of **K** will become increasingly rank deficient. This increasing rank deficiency will exacerbate the limitations of these metrics we document above.

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 shape change. Calculation of this metric would allow us to state definitively that the *Smilodon* data is tighter in shape dispersion than *Canis dirus*, assuming the observed pattern from 14 landmarks holds up. Yet given the 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*. This adaptive difference is easily stated in qualitative terms, but we quantify it here for the first time.

Conclusions

-- Measures of whole-shape integration are reviewed; 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.

-- Dire wolf jaws from two populations from RLB are analyzed for shape and size differences. The two populations come from Pit 13, deposited circa 19 kya, and Pit 61/67, deposited circa 14 kya. Pit 13 wolves show elevated tooth breakage and wear.

-- Pit 13 wolves are significantly smaller than Pit 61/67 wolves. Late stage growth of the jaw corpus relative to the cheek teeth is a prominent feature of dire wolf covariance, as it is in the gray wolf, yet this factor does not differentiate between pits. Neotenic shape change due to retardation of ontogenetic allometry is therefore not the primary driver of shape change between pits (contra O'Keefe et al., 2014).

-- The primary shape difference between pits is canine size and length of the associated condyloid process, here interpreted as a manifestation of size selection down the (strongly positive) static allometry gradient of this complex.

-- The only truly significant module shown in the jaw is the ontogenetic one 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.

-- Classical metrics of whole shape integration fail to recover the evolutionary increase in modularity in Pit 61/67 wolves. They fail because 1) they rely on the correlation matrix, which inflates trivial variance in geometric morphometric data; and 2) they measure only variance dispersion, while ignoring dimensionality dispersion.

-- New metrics based on the Shannon entropy-modified SGV are defined to quantify aspects of phenotypic shapes 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 most general metric treated in this paper; it is formally defined as:

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

-- New whole 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 Materials

Data Availability Statement: The data used in this paper are available as supplementary materials, and are also posted to the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.d7wm37q01

Acknowledgements

We thank the curators and staff at the Tar Pit Museum, specifically 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 also thank Dr. H. David Sheets for his assistance using the IMP programs, including pcagen_7a. We thank the Graduate College for the Marshall University Summer Thesis Award and the Marshall University Drinko Research Fellowship to FRO for funding towards this project. Work by FRO on this project was funded in part by NSF EAR-SGP 1757236 (Project SABER), and by a Drinko Distinguished Research Fellowship to FRO. We grudgingly thank Thom Yorke.

Literature Cited

Adams, D. C. 2016. Evaluating modularity in morphometric data: challenges with the RV coefficient and a new test measure. *Methods in Ecology and Evolution*, *7*(5), 565-572.

Adams, D., Collyer, M., and 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. Journal of Vertebrate Paleontology 22(2): 423-428.

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

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

Brannick, A.L., Meachen, J.A., and O'Keefe, F.R., 2015. Microevolution of jaw shape in the dire wolf, Canis dirus, at Rancho La Brea. La Brea and beyond: The paleontology of asphalt-preserved biotas, ed. JM Harris. Natural History Museum of Los Angeles County, Science Series, 42: 23-32.

Bunea, F., She, Y., and Wegkamp, M. H. 2011. Optimal selection of reduced rank estimators of high-dimensional matrices. The Annals of Statistics, *39*(2), 1282-1309.

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. American Zoologist, *36*(1), 44-50.

Curth, S., Fischer, M. S., and 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.

Drake, A. G., and Klingenberg, C. P. 2010. Large-scale diversification of skull shape in domestic dogs: disparity and modularity. The American Naturalist, 175(3), 289-301.

Fuller, B.T., Fahrni, S.M., Harris, J.M., Farrell, A.B., Coltrain, J.B., Gerhart, L.M., Ward, J.K., Taylor, R.E. and Southon, J.R., 2014. Ultrafiltration for asphalt removal from bone collagen for radiocarbon dating and isotopic analysis of Pleistocene fauna at the tar pits of Rancho La Brea, Los Angeles, California. Quaternary Geochronology 22: 85-98.

Goodrich, E.S. 1930. Studies on the Structure and Development of Vertebrates. Macmillan and Co., London.

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. Proceedings of the National Academy of Sciences 112(16): 4891-4896.

Goswami, A. and Polly, P.D., 2010. Methods for studying morphological integration and modularity. The Paleontological Society Papers 16: 213-243.

Hine, E., and Blows, M. W. 2006. Determining the effective dimensionality of the genetic variance–covariance matrix. Genetics 173(2), 1135-1144.

Jolicoeur, P. 1963. 193. Note: the multivariate generalization of the allometry equation. Biometrics, 19(3): 497-499.

Klingenberg, C. P. 1996. Multivariate allometry. In *Advances in morphometrics* (pp. 23-49). Springer, Boston, MA.

Klingenberg, C. P. 2008. Morphological integration and developmental modularity. Annual Review of Ecology, Evolution and Systematics, 39: 115–132.

Klingenberg CP. 2010. Evolution and development of shape: integrating quantitative approaches. Nature Reviews Genetics 11(9): 623-35.

Klingenberg CP. 2013. Cranial integration and modularity: insights into evolution and development from morphometric data. Hystrix, the Italian Journal of Mammalogy 24(1): 43-58.

Lande, R. 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. Evolution 34(2): 292-305.

Lele, S. and Richtsmeier, J.T. 1991. Euclidean distance matrix analysis: A coordinate-free approach for comparing biological shapes using landmark data. American Journal of Physical Anthropology, 86, 415-427.

Marcus, L.F., 1990. Traditional morphometrics. In Proceedings of the Michigan morphometrics workshop (Vol. 2, pp. 77-122). Ann Arbor, Michigan: The University of Michigan Museum of Zoology.

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. Biological Journal of the Linnean Society 126(4): 721-742.

Meachen, J.A., O'Keefe, F.R., and Sadleir, R.W. 2014. Evolution in the sabre-tooth cat, *Smilodon fatalis*, in response to Pleistocene climate change. Journal of Evolutionary Biology 27: 714-723.

Melo, D. and Marroig, G., 2015. Directional selection can drive the evolution of modularity in complex traits. Proceedings of the National Academy of Sciences, 112: 470-475.

Meloro, C. 2012. Mandibular shape correlates of tooth fracture in extant Carnivora: implications to inferring feeding behavior of Pleistocene predators. Biological Journal of the Linnean Society 106: 70-80.

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., Fet, E. V., & Harris, J. M. 2009. Compilation, calibration and synthesis of faunal and floral radiocarbon dates, Rancho La Brea, California. Contributions in Science 518: 1-16.

O'Keefe, F. R., Binder, W. J., Frost, S. R., Sadlier, R. W. and 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. Palaeontologia Electronica.

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. Journal of Mammalogy 94(6):1223-36.

Pavlicev, M., Cheverud, J. M., & Wagner, G. P. 2009. Measuring morphological integration using eigenvalue variance. Evolutionary Biology 36(1): 157-170.

R Core Team, 2014. R: A language and environment for statistical computing.

Rohlf, F., 2013. tpsDig2, Stony Brook: State University of New York.

Segura, V., Cassini, G. H., Prevosti, F. J., & Machado, F. A., 2020. Integration or Modularity in the Mandible of Canids (Carnivora: Canidae): a Geometric Morphometric Approach. Journal of Mammalian Evolution doi.org/10.1007/s10914-020-09502-z.

Stock, C. and Harris, J.M., 1992. Rancho La Brea: A Record of Pleistocene Life in California. Science Series No. 37. Natural History Museum of Los Angeles County, Los Angeles, CA.

Van Valen, L. (1974). Multivariate structural statistics in natural history. Journal of Theoretical Biology, 45(1), 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 frac- ture is Pleistocene and recent carnivores. Biological Journal of the Linnean Society 96:68-81.

VanValkenburgh, B. and Hertel, F., 1993. Tough times at La Brea: tooth breakage in large carnivores of the late Pleistocene. Science, 261: 456-459.

Wagner, G. P. (1984). On the eigenvalue distribution of genetic and phenotypic dispersion matrices: evidence for a nonrandom organization of quantitative character variation. Journal of Mathematical Biology, 21: 77-95.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. Biometrika: 471-494.

Appendix

R Code for Effective Dispersion Metrics

```
#attach data manipulation package dplyr
library(dplyr)

library(broom)

library(readxl)

#Define a variable containing the X,Y landmark data.The data is read in as a data frame
from an Excel file.

GPALM <- read_excel("~/Desktop/C Dirus 14 LM/CDirus13LM.xlsx")
View(GPALM)

#initialize variables
EffRank=matrix(nrow=1,ncol=0)
SGV=matrix(nrow=1,ncol=0)
EffRankSGV=matrix(nrow=1,ncol=0)
EigenSumRec=matrix(nrow=1,ncol=0)
EffDispersion=matrix(nrow=1,ncol=0)
AdaptDia=matrix(nrow=1,ncol=0)

#Start bootstrap loop
for (i in 1:10000){
        bootsample=sample_n(GPALM, 36, replace = FALSE)
                                        #edit for n and T/F

#Define variable that is the covariance matrix of the above. Be sure to specify only the
landmark columns in the data frame.
CVLM= cov(bootsample[3:30])

#Define object that is the eigenanalysis of the covariance matrix.
CVLMeigen=eigen(CVLM)

EigenVal<-CVLMeigen$values

#Define the summation of eigenvalues from the eigenanalysis.
EigenSum=sum(EigenVal)

#Create vector of scaled eigenvalues
PofK<-EigenVal/EigenSum

#Creat vector of lnPofK and calculate effective rank
```

```
lnPofK<-log(PofK)

Product<-lnPofK*PofK

SumProduct<-sum(Product, na.rm=TRUE)

ShannonEntropy <- -1*SumProduct

EffectiveRank<-exp(ShannonEntropy)

#Begin SGV24 calculation
EigenProduct<-prod(EigenVal[1:24])

StandGenVar<-(EigenProduct)^(1/24)

SGV<-cbind(SGV, StandGenVar)

#Begin ReSGV calculation

ERInteger<-as.integer(EffectiveRank)

EigenTrim<-EigenVal[1:ERInteger]

FractRankVar<-EigenVal[ERInteger+1]*(EffectiveRank-ERInteger)

ProductEffRankEigen<-prod(EigenTrim)*FractRankVar

EffectiveRankSGV<-(ProductEffRankEigen)^(1/EffectiveRank)

#Make Effective Dispersion

EffDisperse<-sqrt(EffectiveRankSGV)

#Divide Effective Dispersion by the Re mean eigenvalue

EffDisHat<-EffDisperse/(sum(EigenTrim)+(EigenVal[ERInteger+1])*(EffectiveRank-
ERInteger))

EffRank <-cbind(EffRank, EffectiveRank)

EffRankSGV <-cbind(EffRankSGV, EffectiveRankSGV)

EigenSumRec <-cbind(EigenSumRec, EigenSum)

EffDispersion <-cbind(EffDispersion, EffDisperse)
```

```
AdaptDia <- cbind(AdaptDia, EffDisHat)

cat("\014")

print(i)

}

#calculate and print st deviation and mean for each parameter

print(mean(EigenSumRec))
print (sd(EigenSumRec))

print(mean(EffRank))
print (sd(EffRank))

print (mean(SGV))
print (sd(SGV))

print (mean(EffRankSGV))
print (sd(EffRankSGV))

print (mean(EffDispersion))
print (sd(EffDispersion))

print (mean(AdaptDia))
print (sd(AdaptDia))
```