Detecting (non)parallel evolution in multidimensional spaces: angles, correlations, and eigenanalysis

Junya Watanabe
Department of Earth Sciences, University of Cambridge,
Downing Street, Cambridge, CB2 3EQ, United Kingdom
jw2098@cam.ac.uk
https://orcid.org/0000-0002-9810-5286

8th October 2021*
doi:10.32942/osf.io/2gxwb

Abstract

Parallelism between evolutionary trajectories in a trait space is often seen as evidence for repeatability of phenotypic evolution, and angles between trajectories play a pivotal role in the analysis of parallelism. However, many biologists have been ignorant on properties of angles in multidimensional spaces, and unsound uses of angles are common in the biological literature. To remedy this situation, this study provides a brief overview on geometric and statistical aspects of angles in multidimensional spaces. Under the null hypothesis that trajectory vectors have no preferred directions, the angle between two independent vectors is concentrated around the right angle, with a more pronounced peak in a higher-dimensional space. This probability distribution is closely related to t- and beta distributions, which can be used for testing the null hypothesis concerning a pair of trajectories. A recently proposed method with eigenanalysis of a vector correlation matrix essentially boils down to the test of no correlation or concentration of multiple vectors, for which a simple test procedure is available in the statistical literature. Concentration of vectors can also be examined by tools of directional statistics such as the Rayleigh test. These frameworks provide biologists with baselines to make statistically justified inferences for (non)parallel evolution.
Keywords: allometric space; directional statistics; high-dimensional data; parallel evolution; phenotypic trajectory analysis.

1 Introduction

Multivariate approaches have proven to be powerful means to analyse phenotypes, yielding more holistic and nuanced understanding of organismal evolution and development than achievable from univariate approaches. It is now fairly common to conceptualise and analyse patterns of phenotypic evolution in multidimensional trait spaces (e.g., Stayton, 2008, 2015; Adams & Collyer, 2009, 2019; Arbuckle *et al.*, 2014; Speed & Arbuckle, 2017; Bolnick *et al.*, 2018). However, increasing dimensionality sometimes poses challenges in interpreting and analysing quantities that superficially appeared familiar. This brief review concerns technical (rather than biological)

*This version (version 3 on *EcoEvoRxiv*) is essentially identical to the 30th September version (version 2 on *EcoEvoRxiv*), except for Figure 2, which was not rendered as intended in the previous version.

aspects of the analysis of phenotypic trajectories in multidimensional spaces, with a particular focus on the angles and their applications to detection of parallel evolution.

A variety of toolkits exists for analysing evolutionary or developmental trajectories in multidimensional spaces. Historically, a notable breakthrough was Jolicoeur's (1963b; 1963a) generalisation of the allometry formula to multivariate cases, which extended bivariate line-fitting into principal component analysis (PCA) (see also Shea, 1985; Klingenberg, 1996, 2016). Motivation has naturally arisen for comparing patterns of multivariate allometric variation between taxa or populations (e.g., Zelditch *et al.*, 2003, 2016; Urošević *et al.*, 2013; Wilson, 2013, 2018; Rohner, 2020; Pérez-Ben *et al.*, 2020; Feiner *et al.*, 2021). One useful concept is the allometric space (Solignac *et al.*, 1990; Klingenberg & Froese, 1991; Gerber *et al.*, 2008), in which various visualisations and analyses are possible via ordination, clustering, etc., by treating empirical allometric axes as observations (Gerber *et al.*, 2008; Baliga & Mehta, 2018; Strelin *et al.*, 2018). Another broadly employed tool is the phenotypic trajectory analysis (Adams & Collyer, 2007, 2009; Collyer & Adams, 2007, 2013), which primarily concerns quantification and statistical testing of inter-population differences in the magnitude, direction, and shape of phenotypic trajectories in a trait space.

Recently, the phenotypic trajectory analysis, with the name of phenotypic change vector analysis, has fuelled investigations into the parallel evolution (Oke *et al.*, 2017; Bolnick *et al.*, 2018). Here, the term parallel evolution is used in the geometric sense; parallelism between trajectories in a trait space between multiple ancestor–descendant pairs (Stayton, 2006; Bolnick *et al.*, 2018), which typically results in acquisition of similar derived traits in the descendants. Parallel responses to similar selection pressures between lineages are often regarded as evidence for repeatability or predictability of phenotypic evolution under natural selection, but the prevalence and extent of such parallelism are subjects of ongoing debate (e.g., Agrawal, 2017; Blount *et al.*, 2018; Rincon-Sandoval *et al.*, 2020). A recent trend is to quantitatively analyse patterns of evolutionary changes in what are regarded as typical examples of parallel evolution (e.g., Kaeuffer *et al.*, 2012; Raeymaekers *et al.*, 2017; Langerhans, 2018). The angles between phenotypic change vectors of different lineages play an especially pivotal role in empirical analyses of parallel evolution (e.g., Stuart *et al.*, 2017; Oke *et al.*, 2017; Haines *et al.*, 2020; James *et al.*, 2021; Owens *et al.*, 2021; Weber *et al.*, 2021), because they are supposed to provide 'intuitive and mathematically formal' measures of (non)parallelism (Stuart *et al.*, 2017: p. 6).

Unfortunately, however, interpretation of angles in multidimensional spaces is not as straightforward as some biologists have assumed. Consider, for example, the angle between randomly directed vectors in 2- and 3-dimensional spaces. It is convenient to fix one of them pointing an arbitrary 'pole' and to let the other be uniformly distributed on the unit circle and sphere (Fig. 1A, B). The probability density of the angle between these vectors is then proportional to the arc length and surface area for a given infinitesimal increment of 'latitude'. One will notice that the density for the 2-dimensional space is uniform (Fig. 1A), whereas that for the 3-dimensional space is peaked at the 'equator' because this region encompasses more area per latitude than 'polar' regions (Fig. 1B). This simple example demonstrates that distributions of random angles depend on the dimensionality, warning against extending our naïve intuition into high-dimensional spaces. Regrettably, few recent biologists studying evolutionary parallelism

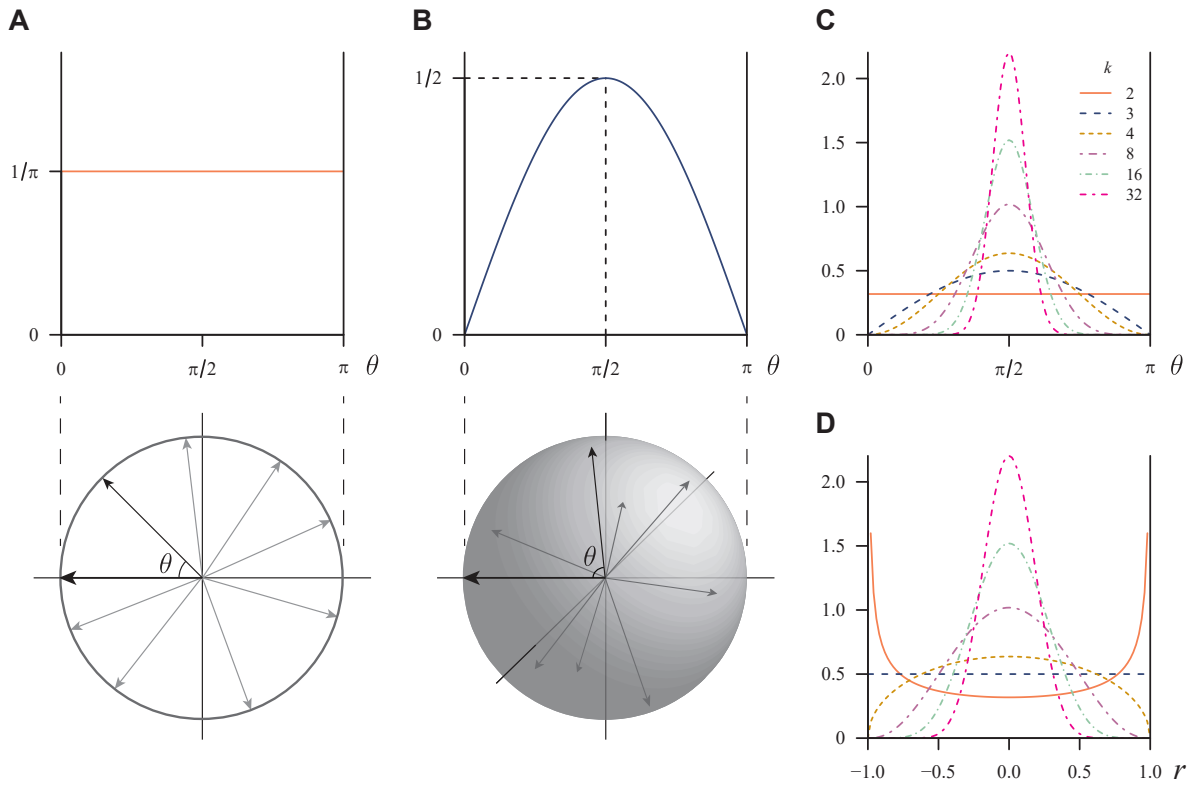

Figure 1. Distribution of angle in multidimensional spaces. **A, B**) Probability density of angle θ between two vectors uniformly distributed on 2-dimensional circle and 3-dimensional sphere, respectively. Lower panels show schematic illustrations of the angle between the vector pointing a 'pole' (the thick arrow pointing the left-hand side) and another uniformly distributed on the unit circle/sphere. Upper panels show the corresponding densities. **C**) Density of θ in general k-dimensional cases (eq. 9). **D**) Density of $r = \cos\theta$ (eq. 8).

have appropriately taken this trend into account. Frameworks to make statistically justified inferences on angles have essentially been lacking in the current empirical literature.

This paper gives a brief overview on methods to analyse angles in multidimensional spaces. Specifically, it first derives the probability distribution of the angle between random vectors under the null hypothesis that the vectors do not have preferred directions. It is by no means novel to science or even to the biological literature, where relevant results have been used in one form or another (e.g., Rice, 1990; Klingenberg & Marugán-Lobón, 2013; Ram & Hadany, 2015; Thompson *et al.*, 2019). The primary aim here is to disseminate well-known results with theoretical underpinnings.

Recently, De Lisle & Bolnick (2020) proposed a framework for analysing multiple vectors simultaneously via eigenanalysis of a vector correlation matrix. Although elegant in design, this framework lacked clear justifications as to which summary statistic should be looked at. This study gives an alternative interpretation and a simple test statistic for this framework. Potentially useful exploratory and visualisation methods for phenotypic change vectors are also discussed in connection to the eigenanalysis.

2 Theory

2.1 Preliminaries

To begin with, let us review the definition of the ordinary (Pearson product-moment) correlation coefficient, which has a close relationship with angles between random vectors. For the bivariate random observations of size N, $(x_1, x_2, \ldots, x_N)$ and $(y_1, y_2, \ldots, y_N)$, the correlation coefficient r is defined as

$$r = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \sum_{i=1}^{N} (y_i - \bar{y})^2}}, \tag{1}$$

where $\bar{x}$ and $\bar{y}$ are the sample means: $\bar{x} = \sum_{i=1}^{N} x_i/N, \bar{y} = \sum_{i=1}^{N} y_i/N$. It is elementary that r ranges between -1 and 1, with values closer to 0 suggesting no correlation between the variables. By using the matrix notation $\mathbf{x} = (x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_N - \bar{x})^T$ and $\mathbf{y} = (y_1 - \bar{y}, y_2 - \bar{y}, \ldots, y_N - \bar{y})^T$, where the superscript ^{T} denotes transpose, we can rewrite equation 1 as

$$r = \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}, \tag{2}$$

where the numerator is the inner product, and $\|\cdot\|$ denotes the vector norm or length ($\|\mathbf{x}\| = \sqrt{\mathbf{x}^T \mathbf{x}}$). Recall the geometric definition of the inner product,

$$\mathbf{x}^T \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos\theta, \tag{3}$$

where θ is the angle formed by $\mathbf{x}$ and $\mathbf{y}$ in their N-dimensional space. Then, we have

$$r = \cos\theta. \tag{4}$$

That is, the correlation coefficient and the angle between a pair of random vectors are directly related through the cosine/arccosine transformation. Here, the range of θ is taken as $[0, \pi]$ (in radians) so that a one-to-one, though negative, relationship exists between r and θ: in the case of perfect positive correlation, $r = 1$, the two vectors point to the same direction, $\theta = 0$; in the case of no correlation, $r = 0$, and the two vectors are perpendicular to each other, $\theta = \pi/2$.

We could standardise the variables by their standard deviations before calculating r. That is, by putting $\mathbf{u} = \|\mathbf{x}\|^{-1}\mathbf{x}$ and $\mathbf{v} = \|\mathbf{y}\|^{-1}\mathbf{y}$, we could simplify the notation

$$r = \cos\theta = \mathbf{u}^T \mathbf{v}. \tag{5}$$

Since $\|\mathbf{u}\| = \|\mathbf{v}\| = 1$, $\mathbf{u}$ and $\mathbf{v}$ denote points on the unit hypersphere in the N-dimensional space.

Technically, the sample-mean-centred vectors $\mathbf{x}$ and $\mathbf{y}$ are in an $(N-1)$-dimensional space, because centring with the sample mean reduces the effective dimensionality—the so-called degree of freedom—of the original N-vectors by one. It is well established that, for normal (and other) variables, the distribution of r with N sample-mean-centred observations from a population with arbitrary mean is the same as that with $N-1$ observations centred at a known population mean (e.g., Hotelling, 1953; Anderson, 2003). For what follows, it is convenient to consider the latter with the population mean 0 (see below).

This discussion concerns the equivalence between correlations in the trait space and angles in the sample (lineage, species, etc.) space, but the same relationship also holds when the space labels are swapped; i.e., the equivalence between correlations in a sample space and angles in a trait space. We now turn to the distribution of random angles with a general k-dimensional space.

2.2 Distribution of random angles

Let us consider a pair of random vectors $\mathbf{x} = (x_1, x_2, \ldots, x_k)^T$ and $\mathbf{y} = (y_1, y_2, \ldots, y_k)^T$ and the angle θ between them. The elements are assumed to be independently and identically distributed. Let $b = (\mathbf{y}^T\mathbf{x})/(\mathbf{x}^T\mathbf{x})$, the ratio of the inner product between $\mathbf{x}$ and $\mathbf{y}$ to the squared norm of $\mathbf{x}$. By the geometric definition of the inner product (eq. 3), the vector $b\mathbf{x}$ points to the foot of the perpendicular from $\mathbf{y}$ to $\mathbf{x}$, and the vector $\mathbf{y} - b\mathbf{x}$ denotes this perpendicular (Fig. 2). In the terminology of regression, $b\mathbf{x}$ and $\mathbf{y} - b\mathbf{x}$ are predictions and residuals, respectively, in the regression of $\mathbf{y}$ on $\mathbf{x}$ (without intercept). The angle θ is related to these vectors in the trigonometric relationship

$$\frac{1}{\tan\theta} = \frac{b\|\mathbf{x}\|}{\|\mathbf{y} - b\mathbf{x}\|}. \tag{6}$$

The distribution of this quantity is heuristically derived here; see, e.g., Hotelling (1953), Muirhead (1982: section 5.1), Anderson (2003: section 4.3), or Cai *et al.* (2013) for formal proofs. Assume the null hypothesis that the elements of $\mathbf{x}$ and $\mathbf{y}$ are normally distributed with mean 0 and variance σ^2, and that these two vectors are independent. The standardised vectors $\|\mathbf{x}\|^{-1}\mathbf{x}$ and $\|\mathbf{y}\|^{-1}\mathbf{y}$ are uniformly distributed on the unit hypersphere in the k-dimensional space. We can rotate the coordinate axes arbitrarily as far as concerning the distribution of θ; let $\|\mathbf{x}\|^{-1}\mathbf{x} = (1, 0, \ldots, 0)^T$ for simplicity. Then, the distribution of $b\|\mathbf{x}\| = \mathbf{y}^T(\|\mathbf{x}\|^{-1}\mathbf{x}) = y_1$ is normal with mean 0 and variance σ^2. Also, that of $\|\mathbf{y} - b\mathbf{x}\|^2/\sigma^2 = \sum_{i=2}^{k} y_i^2/\sigma^2$ is chi-square with $k - 1$ degrees of freedom, and independent of $b\|\mathbf{x}\|$ (Fig. 2). Therefore, by the operational definition of the t-distribution—namely, the distribution of the ratio of a standard normal variate to the square root of a chi-square variate divided by its degree of freedom, with the two variates

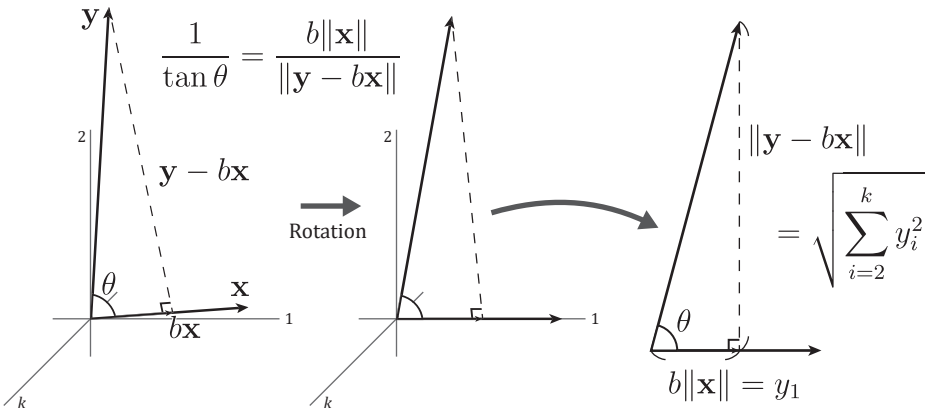

Figure 2. Scheme to find probability distribution of random angle in k-dimensional space (only the 1st, 2nd, and kth coordinate axes are shown for obvious visual restrictions). See text for details.

independent from each other—the quantity

$$\frac{b\|\mathbf{x}\|/\sigma}{\sqrt{\|\mathbf{y}-b\mathbf{x}\|^2/\sigma^2(k-1)}} = \frac{\sqrt{k-1}}{\tan\theta} = \sqrt{k-1}\frac{r}{\sqrt{1-r^2}} \tag{7}$$

has a t-distribution with $k-1$ degrees of freedom. The probability density (or mass, to be strict) of r in this case can be derived by transforming that of the t-distribution:

$$\frac{1}{B[1/2,(k-1)/2]}(1-r^2)^{\frac{k-3}{2}}\,\mathrm{d}r, \quad -1 \le r \le 1, \tag{8}$$

where $B(a,b)$ is the beta function with the two parameters a and b (this is just a normalising constant, whose actual value needs not concern most readers) (see Fig. 1D). Then the density for $\theta = \arccos r$ is, by noting $|\mathrm{d}r| = |-\sin\theta\,\mathrm{d}\theta|$,

$$\frac{1}{B[1/2,(k-1)/2]}(1-\cos^2\theta)^{\frac{k-3}{2}}|-\sin\theta\,\mathrm{d}\theta| = \frac{1}{B[1/2,(k-1)/2]}\sin^{k-2}\theta\,\mathrm{d}\theta, \quad 0 \le \theta \le \pi. \tag{9}$$

This density has a peak at $\theta = \pi/2$, which is increasingly pronounced as k increases (Fig. 1C). Another useful expression can be derived for $s = r^2$, by noting $|\mathrm{d}r| = |\mathrm{d}s/2\sqrt{s}|$ and duplication of the positive and negative branches for r in equation 8:

$$\frac{1}{B[1/2,(k-1)/2]}s^{-\frac{1}{2}}(1-s)^{\frac{k-3}{2}}\,\mathrm{d}s, \quad 0 \le s \le 1, \tag{10}$$

which is the density of the beta distribution with the parameters $1/2$ and $(k-1)/2$.

The same distribution can be obtained from looser conditions than assumed here. For example, $\mathbf{x}$ could be from any distribution (or a fixed vector) as long as it is independent of $\mathbf{y}$ that in turn has a spherically contoured distribution (Muirhead, 1982; Anderson, 2003). Indeed, expressions equivalent to equations 9 and 10 can be obtained from purely geometric evaluation of the surface area of a hyperspherical cap (Rice, 1990; Li, 2011), which is equivalent to the probability for a random vector uniformly distributed on the hypersphere to fall within the region (see also Ram & Hadany, 2015). A similar geometric reasoning was in fact involved in Fisher's (1915; 1925) formal derivation of the t-distribution (see also Stuart & Ord, 1994: chapter 11), so, to be strict, the above derivation was partly circular.

These results can be used for testing the null hypothesis that two phenotypic change vectors have no preferred directions (population means being $(0,\ldots,0)^T$) and are independent from each other, by inserting the dimensionality of the trait space into k. In particular, the P-value for an observed angle can be calculated from the t statistic (eq. 7); example functions for the `R` environment (R Core Team, 2019) are provided in Supplemental Material. This is equivalent to the ordinary correlation test, where typically $k = N-1$ (see above). When the polarities of the vectors are to be ignored (e.g., test for angles between eigenvectors), the beta distribution (eq. 10) can be used instead. An equivalent test is commonly used for testing differences between allometric axes (e.g., Klingenberg & Marugán-Lobón, 2013).

2.3 Pairwise angles and correlations

The above results concern a pair of random vectors, which should suffice when there are only a few lineages to compare. In practice, interest is often in analysing a set of many lineages simultaneously (e.g., Stuart *et al.*, 2017; De Lisle & Bolnick, 2020; Owens *et al.*, 2021). A convenient way to summarise the overall pattern is to construct a matrix of pairwise angles or correlations. Let $\mathbf{x}_i$ denote phenotypic change vectors of p traits from n lineages ($i = 1, \ldots, n$), each starting from its respective ancestor, and arrange these in rows of the $n \times p$ matrix $\mathbf{X}$. This matrix then is standardised so that each row has the length of unity:

$$\mathbf{Z} = \mathrm{diag}(\|\mathbf{x}_i\|^{-1})\mathbf{X}, \tag{11}$$

where $\mathrm{diag}(\cdot)$ denotes an $n \times n$ diagonal matrix with the designated ith diagonal elements. Then we consider the following $n \times n$ inter-lineage correlation matrix

$$\mathbf{C} = \mathbf{Z}\mathbf{Z}^T. \tag{12}$$

By construction, $\mathbf{C}$ is symmetric and its (i, j)-th elements are the vector correlations between the ith and jth vectors (eq. 5), with the diagonal elements being 1. The rows need not be centred, thus retain the full effective dimensionality of p, unless the traits themselves are linearly dependent (as is the case for shape variables; see below). Taking element-wise arccosines of $\mathbf{C}$ yields a matrix of pairwise angles. For sake of discussion, let $\mathbf{\Gamma}$ be the population (true) correlation matrix corresponding to $\mathbf{C}$.

It might be tempting to make statistical inferences by treating pairwise angles or correlations in these matrices as a sample; e.g., calculating mean and standard deviation from all pairwise angles and conducting a t-test for the difference of the mean from, say, $\pi/2$ to detect a parallel signal (Owens *et al.*, 2021). However, such inferences should be, if at all, made with extreme caution, because pairwise angles and correlations are generally not independent from one another. The ordinary t-test and the like assume the observations to be independent (or at least uncorrelated), and violation of this assumption leads to suboptimal performance, e.g., inflated type I error rates. Off-diagonal elements of $\mathbf{C}$ have nonzero covariances unless $\mathbf{\Gamma} = \mathbf{I}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix (Olkin & Siotani, 1976; Olkin & Finn, 1990). Similar should be the case for pairwise angles. Therefore, it is inadvisable to conduct tests for pairwise correlations or angles in this way, unless, perhaps, the covariances are appropriately taken into account (methods for which are available for correlations; Olkin & Finn, 1995; Zou, 2007). On the other hand, it would be valid to conduct a sensibly constructed Monte Carlo test. That said, it is rather questionable whether this test is of any practical use. There are more straightforward ways to test the null hypothesis $\mathbf{\Gamma} = \mathbf{I}_n$ (below), and other cases hardly translate into particular values of mean of pairwise angles or correlations.

2.4 Eigenanalysis and one-step test for multiple vectors

De Lisle & Bolnick (2020) proposed to use eigenanalysis of the inter-lineage correlation matrix $\mathbf{C}$ to detect concentration of phenotypic change vectors in a trait space. That is, to consider

spectral decomposition (or eigendecomposition) of $\mathbf{C}$:

$$\mathbf{C} = \mathbf{ULU}^T, \tag{13}$$

where $\mathbf{U}$ is an $n \times n$ matrix of eigenvectors, and $\mathbf{L} = \text{diag}(l_i)$ is an $n \times n$ diagonal matrix of eigenvalues. Their motivation was to quantify the magnitude of parallelism and effective dimensionality of parallel trajectories in the trait space by analysing eigenvalues of $\mathbf{C}$, which represent variances along the corresponding principal components (PCs). For those purposes, however, it is more straightforward to consider the $p \times p$ inter-trait cross-product matrix $\mathbf{A}$ and its eigendecomposition instead:

$$\mathbf{A} = \mathbf{Z}^T\mathbf{Z} = \mathbf{VKV}^T, \tag{14}$$

where $\mathbf{V}$ is a $p \times p$ matrix of eigenvectors, and $\mathbf{K} = \text{diag}(k_i)$ is a $p \times p$ diagonal matrix of eigenvalues. The non-zero eigenvalues of $\mathbf{C}$ and $\mathbf{A}$ are in fact identical (Appendix A). $\mathbf{C}$ provides a quick means to surmise closeness between phenotypic change vectors, as well as a useful test described below. However, concerning variation in the trait space, $\mathbf{V}$ and $\mathbf{K}$ are more interpretable than $\mathbf{U}$ and $\mathbf{L}$ because the former pair pertains to the p-dimensional trait space whereas the latter pertains to the n-dimensional lineage space (Appendix A). The rest of this study addresses quantification and test of the magnitude of parallelism—the first objective of the eigenanalysis as proposed by De Lisle & Bolnick (2020). Their second objective—determination of dimensionality of parallel trajectories—requires more methodological clarification and elaboration than can be covered here (see Appendix A for comments).

Concentration of variation in a trait space is often measured by dispersion of eigenvalues (Cheverud *et al.*, 1983; Wagner, 1984; Pavlicev *et al.*, 2009a; Haber, 2011; Watanabe, 2021), which is easier to quantify than 'skewness' mentioned (but not quantified) by De Lisle & Bolnick (2020). If phenotypic change vectors are uniformly directed in the trait space, eigenvalues of $\mathbf{A}$ (or $\mathbf{C}$) exhibit low dispersion. If the vectors are concentrated in certain directions, then the eigenvalues are highly dispersed. One complexity here is the presence of sampling error and bias, which render sample eigenvalues more dispersed than the corresponding population eigenvalues (e.g., Anderson, 1963; Jolliffe, 2002; Watanabe, 2021). De Lisle & Bolnick (2020) suggested to compare eigenvalues of $\mathbf{C}$ with Monte Carlo distributions of eigenvalues of matrices drawn from a Wishart distribution. However, that distribution pertains to unscaled cross-product or covariance matrices, so is not directly applicable to a correlation matrix like $\mathbf{C}$, except as a rough approximation. That method would work if the random matrices are also scaled as correlation matrices. Nevertheless, concerning detection of concentration between vectors, there is a much simpler alternative.

It is possible to show that dispersions of eigenvalues of $\mathbf{C}$ and $\mathbf{A}$ (denoted l_i and k_i, respectively) are equivalent to sum of squared correlations (see Appendix A):

$$\sum_{i=1}^{n}(l_i - \bar{l})^2 = \sum_{i=1}^{p}(k_i - \bar{k})^2 + \frac{n^2}{p} - n = 2\sum_{i<j}^{n} r_{ij}^2, \tag{15}$$

where $\bar{l}$ and $\bar{k}$ are the averages of eigenvalues, and r_{ij} are the (i, j)-th elements of $\mathbf{C}$. Therefore, instead of eigenvalue dispersion, we can equivalently consider sum of squared correlations $\sum_{i<j}^{n} r_{ij}^2$,

which is a straightforward measure of deviation from independent directions of vectors. Let us assume the null hypothesis of $\mathbf{\Gamma} = \mathbf{I}_n$; that is, all vectors are independently directed from one another without preferred directions (in the population). For n lineages, we take as if p traits are observations. Under the multivariate normality of the elements of $\mathbf{X}$, each of r_{ij}^2 is distributed as Beta$[1/2, (p-1)/2]$ (eq. 10) and hence has the mean $1/p$ and variance $2(p-1)/p^2(p+2)$. Furthermore, it is possible to show that r_{ij}^2's are uncorrelated with one another under the null hypothesis (Schott, 2005; Watanabe, 2021). Therefore, the expectation and variance of the sum of squared correlations are:

$$\mathrm{E}\left(\sum_{i<j}^{n} r_{ij}^2\right) = \frac{n(n-1)}{2p}, \quad \mathrm{Var}\left(\sum_{i<j}^{n} r_{ij}^2\right) = \frac{n(n-1)(p-1)}{p^2(p+2)}. \tag{16}$$

From these moments, Schott (2005) proposed the following high-dimensional asymptotic test. Under the condition $n \to \infty$, $p \to \infty$, and $n/p \to \gamma \in (0, \infty)$, the distribution of $\sum_{i<j}^{n} r_{ij}^2 - n(n-1)/2p$ converges to the normal distribution with mean 0 and variance $\lim\left[\mathrm{Var}(\sum_{i<j}^{n} r_{ij}^2)\right] = \gamma^2$. (This condition may at first look odd, but is just a modest generalisation from the ordinary large-sample asymptotic condition, $n \to \infty$ and $p/n \to 0$, which is equally unrealistic.) Empirical values of $\sum_{i<j}^{n} r_{ij}^2$ can be compared with the normal distribution with the above mean and variance (eq. 16), and a large deviation can be seen as evidence against the null hypothesis, suggesting concentration of vectors. Schott (2005) showed by simulations that this test has a reasonable type I error rate (although slightly too liberal when p or n is small, e.g., <16) and a power usually superior to that of the conventional likelihood-ratio test.

An obvious caveat on this procedure is that the test statistic does not convey information on the signs of correlation coefficients (neither do the eigenvalues). Therefore, this test does not distinguish parallel and antiparallel signals. At least the original correlation matrix or PC scores should be inspected to surmise what type of deviation from the null is present (De Lisle & Bolnick, 2020).

2.5 Rayleigh test for unimodal concentration

The test of no correlation described above is a generic way to detect concentration of random vectors in any form. If the detection of parallel signal is of particular interest, it is probably more adequate to use the Rayleigh test from directional statistics (Mardia *et al.*, 1979; Mardia & Jupp, 1999). It aims to test whether the sample is from the uniform distribution (the null) or a unimodal distribution (the alternative) on the unit hypersphere. In particular, the von Mises–Fisher distribution—an analogue of the normal distribution for directional data—is assumed as a model.

The Rayleigh test proceeds as follows. Let $\mathbf{z}_i$ denote the rows of $\mathbf{Z}$ defined above (eq. 11); these are p-dimensional vectors for n lineages, standardised to have the length of unity. In addition, define the mean vector as $\bar{\mathbf{z}} = \sum_{i=1}^{n} \mathbf{z}_i/n$. Clearly, $\bar{\mathbf{z}}$ lies within the unit hypersphere as it is the centroid of $\mathbf{z}_i$'s. Under the null hypothesis of uniform distribution, $\bar{\mathbf{z}}$ should lie close to the origin. The more (unimodally) concentrated $\mathbf{z}_i$'s are, the farther apart $\bar{\mathbf{z}}$ lies from the origin. Under the null hypothesis, the expectation and variance of $\mathbf{z}_i$ are $\mathbf{0}$ and $p^{-1}\mathbf{I}_p$, respectively (see

Mardia *et al.*, 1979; Anderson, 2003). It then follows from the classic central limit theorem that $\bar{\mathbf{z}}$ is normally distributed with mean $\mathbf{0}$ and variance $(np)^{-1}\mathbf{I}_p$ as $n \to \infty$ under the null hypothesis. Therefore, it is possible to construct a test based on the following test statistic and the asymptotic null distribution:

$$S = np\|\bar{\mathbf{z}}\|^2 \sim \chi^2_p, \quad n \to \infty, \tag{17}$$

where χ^2_p denotes the chi-square distribution with p degrees of freedom. For a large value of S, the null hypothesis is rejected, suggesting concentration of random vectors. It is recommended to use the following correction, which is known to yield a better approximation of the same limiting distribution (Mardia & Jupp, 1999):

$$S^* = \left(1 - \frac{1}{2n}\right) S + \frac{1}{2n(p+2)} S^2. \tag{18}$$

The Rayleigh test is a particularly powerful test to detect unimodal concentration of random directional vectors (Mardia *et al.*, 1979; Mardia & Jupp, 1999), and is also known to be robust against high p/n ratios (Ley & Verdebout, 2017). However, it is insensitive to such forms of concentration that yield $\bar{\mathbf{z}}$ near the origin. Examples include balanced multimodal concentration (e.g., equal parallel and antiparallel signals) and uniform distribution on a large circle. There exist several other tests of concentration for those cases (Mardia & Jupp, 1999; Cai *et al.*, 2013; Pewsey & García-Portugués, 2021), among which Schott's (2005) test described above could be placed as well.

3 Example analysis

Stuart *et al.*'s (2017) dataset of lake–stream divergence in the threespine stickleback (*Gasterosteus aculeatus*) is re-analysed here for demonstration. The original data were pre-processed as described in Appendix B. The resultant dataset consists of 13 phenotypic change vectors in 80 nominal morphological traits: 41 linear measurements, 38 Procrustes-aligned shape coordinates and 1 centroid size (from 2-dimensional geometric morphometric analysis of 19 full landmarks). The effective dimensionality of the vectors is $80 - 4 = 76$, as 4 degrees of freedom is lost by Procrustes alignment (assuming that the configurations were projected onto the tangent space; when this was not done, complexity would arise because the coordinates are nonlinearly dependent).

The resultant 78 pairwise angles ranged from 0.49 to 2.62 (28.0°–149.9°). Compared with the null distribution of angles in the 76-dimensional space (eq. 9), 38 and 29 out of these were closer to parallel and antiparallel, respectively, than expected from random directions by chance alone (two-sided test at $\alpha = 0.05$; no rigorous error rate control is deemed necessary for this demonstrative analysis; Fig. 3A). The mean angle of 1.50 (86.0°) was closer to parallel than expected for a mean of 78 random angles ($P < 1 \times 10^{-5}$ based on a Monte Carlo simulation with 10^5 iterations; Fig. 3B). This interpretation is in stark contrast with that of Stuart *et al.* (2017), who regarded their mean of 81.1° with 84 traits as 'nearly orthogonal' without considering the dimensionality. Note, however, that this test is for illustrative purposes only, as the mean pairwise angle lacks a clear interpretability (see above).

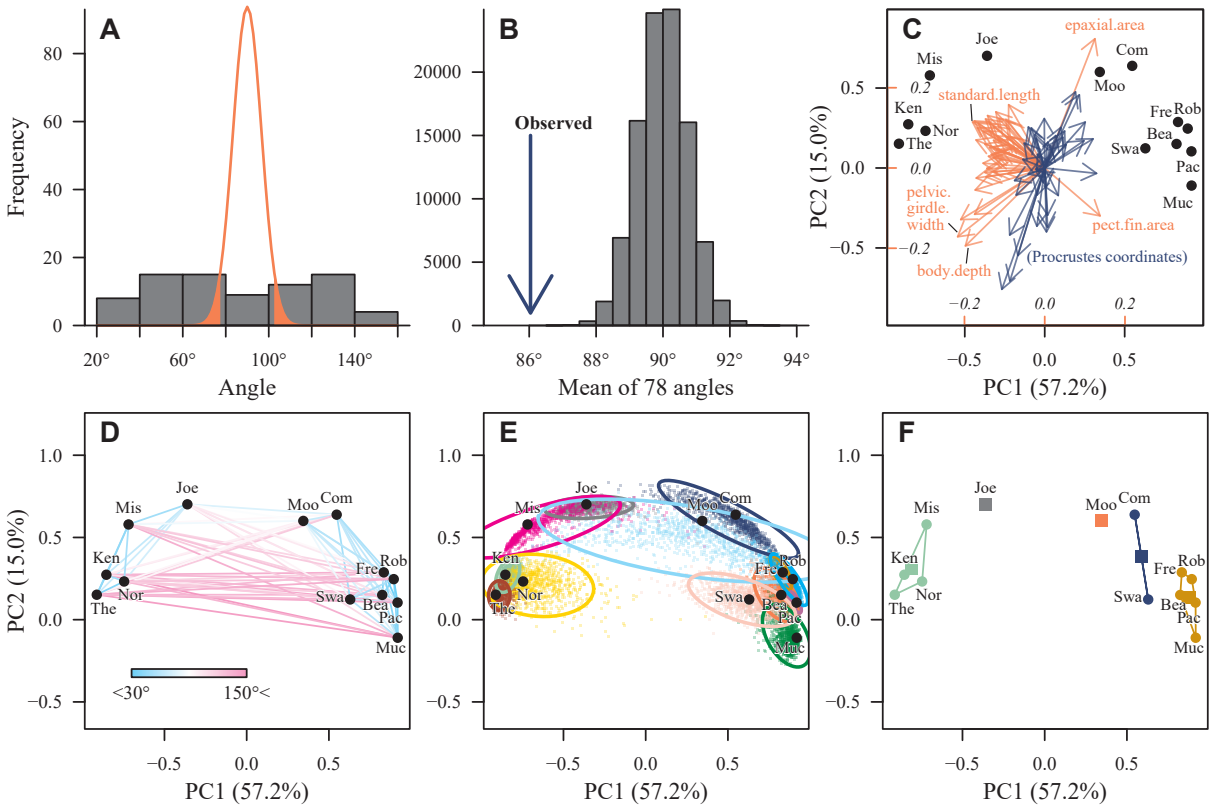

Figure 3. Re-analysis of Stuart *et al.*'s (2017) dataset. **A**) Histogram of 78 pairwise angles between the phenotypic change vectors in 13 lineages of *Gasterosteus aculeatus*, compared with a scaled density of random angles for $p = 76$ (eq. 9). Regions outside the 2.5 and 97.5 percentiles of the density are shown with solid orange fills. **B**) Mean of the 78 pairwise angles (blue arrow) compared with the null distribution (histogram) based on 100,000 Monte Carlo simulation runs. **C**–**F**) Principal component plots of phenotypic change vectors with different visualisations. **C**) PCA biplot showing scores (points) and coefficients (arrows) of PC1 and PC2. The scaling parameter α was set to 1, so that Euclidean distances between observations are maximally preserved (see Appendix C). Blue arrows denote Procrustes shape coordinates, which cannot be interpreted individually, whereas orange ones denote the other traits, some of which are labelled. The inner axis labels are for coefficients, whereas the outer ones are for scores. **D**) Pairwise angles shown with colour-scaled segments. **E**) Clouds of bootstrap replicates and approximate 95% confidence ellipses. Ellipses are based on 5000 replicates of PC scores, but only 1000 replicates are shown for visual clarity. **F**) Grouping with k-means clustering shown with colours and convex hulls. This grouping gave the smallest within-group sum of squares for (arbitrarily chosen value of) $k = 5$. Squares denote group centroids. Note that the clustering was conducted on the full trait space rather than this 2-dimensional plane. Acronyms for watersheds are as in Stuart *et al.* (2017).

Schott's (2005) test described above was applied to the inter-lineage correlation matrix **C**. Sum of the 78 squared correlations was 24.25, whereas the null expectation and standard deviation (eq. 16 with $n = 13$ and $p = 76$) were 1.03 and 0.16, respectively, indicating a statistically significant deviation from the null hypothesis of independent directions ($Z = 144.08$; $P < 1 \times 10^{-10}$). The Rayleigh test also detected a significant concentration ($S^* = 141.49$; 76 degrees of freedom; $P = 0.000008$), although the interpretation is not so obvious in the presence of both parallel and antiparallel signals in the dataset.

Ordination from PCA of the standardised phenotypic change vectors **Z** is shown in Figure 3C–F with different visualisations (see Appendix C for details). PCA biplot shows that the first two PCs are strongly loaded with standard length and other traits highly correlated with it, while additional variation is provided by traits like pelvic girdle width and body depth (Fig. 3C). The

vectors of some lineages appear closely clustered with one another, but distribution of PC scores across the origin indicates that not all lineages had similar divergence (Fig. 3D). Nonparametric bootstrapping suggests that differences between vectors are mostly larger than what would be expected from sampling error alone except in most similar pairs (Fig. 3E). Nevertheless, the magnitude of sampling error appears heterogeneous among lineages, cautioning against face-value interpretation of differences; for example, sampling error is evidently large for the Moore watershed, and this is not due to small sample size (or distortion of the plot). Potential clustering was explored with k-means clustering with varying numbers of clusters, and the result for $k = 5$ is shown as an example (Fig. 3F). The results collectively indicate that both parallel and antiparallel signals are present, as previously found by De Lisle & Bolnick (2020) from their re-analysis of the same dataset.

4 Discussion

Angles have been commonly used in quantitative analyses of parallel evolution, but their properties in multidimensional spaces have not attained due attention from biologists. As clarified by the above analysis, angles between randomly directed vectors is peaked around the right angle in multidimensional spaces (eq. 9; Fig. 1). It is therefore inadvisable to interpret angles at face value, e.g., angles closer to 90° than 0° regarded as evidence against parallel evolution on their own (Stuart *et al.*, 2017; Oke *et al.*, 2017). In addition, the dependency of the peakedness on dimensionality (eq. 9; Fig. 1C) renders angles incomparable across different dimensions. Thus, direct comparison of angles or pooled meta-analysis across varying dimensionalities (Oke *et al.*, 2017; Haines *et al.*, 2020; Radersma *et al.*, 2020) will not be meaningful, unless dimensionality is sensibly taken into account. A potentially useful standardisation in this respect is $\sqrt{k-2}(\pi/2-\theta)$, whose distribution under the null condition (eq. 9) converges to the standard normal distribution as $k \to \infty$ (Cai *et al.*, 2013); when k is sufficiently large, this quantity could be used as an effect size against the null distribution.

This review concentrated on the null hypothesis that vectors are independent and have no preferred directions, which is just one of many hypotheses of potential biological interest (Bolnick *et al.*, 2018; De Lisle & Bolnick, 2020). This is not to claim for superior biological importance of this hypothesis over another, but rather to present it as a baseline for analysing multidimensional vectors. On the other extreme, the hypothesis of completely parallel vectors (in the population) could be tested, if interest is in detecting deviation from parallelism (Bolnick *et al.*, 2018). It is, however, more difficult to define a unified procedure for testing this null hypothesis than it may seem. It should in principle be possible to extend the present parametric framework into any arbitrary population values of correlation (although the distributions are substantially more complex). However, a practical test procedure will need to incorporate sampling error, whose nature and magnitude would largely depend on individual study systems. This is partly because complete (anti)parallelism in the population eliminates any room for sampling variation, thereby trivially yielding sample correlation coefficient exactly 1 or -1 with probability 1. A more realistic option will be to adopt one of the resampling-based approaches (Klingenberg, 1996; Adams & Collyer, 2009; Collyer & Adams, 2013; Sheets & Zelditch, 2013), as is done in the original phenotypic trajectory analysis. However, it should be remembered

that a resampling-based test, although being nominally nonparametric, is usually not free from the assumption that the populations share the same form of distribution, potentially differing only in the quantity to be looked at (e.g., Anderson, 2001; Manly & Navarro Alberto, 2020). Between-group heteroscedasticity, whose presence was also suggested in the present re-analysis (Fig. 3E), can possibly undermine adequacy of tests of this type. Robustness of resampling-based tests against such cases needs to be critically assessed. The literature of directional statistics (e.g., Mardia & Jupp, 1999; Ley & Verdebout, 2017; Pewsey & García-Portugués, 2021) may potentially provide useful directions for alternatives.

Apart from hypothesis testing, exploratory methods such as ordination and clustering can provide useful insights into variation among phenotypic change vectors (Appendix C). It should be straightforward to apply concepts and techniques originally devised for the analysis of allometric space to phenotypic change vectors. Examples include quantification of allometric disparity (e.g., Gerber *et al.*, 2008; Urošević *et al.*, 2013), test for shared trajectories in subspaces (Klingenberg, 1998; Mitteroecker *et al.*, 2005; Gerber *et al.*, 2007), and simultaneous visualisation with phylogeny (Baliga & Mehta, 2018). Although not fully discussed here, clustering approaches may also be useful in detecting and summarising patterns in multiple phenotypic change vectors (Fig. 3F; Appendix C).

A paramount assumption in almost any geometric analysis in a trait space (Adams & Collyer, 2009; Stayton, 2015; Bolnick *et al.*, 2018) is that vectors can be meaningfully compared across different regions of the trait space. This is, for example, when all traits are measured in the same unit and no bounds exist (at least in the region of practical concern). If traits are in different units (e.g., linear measurements and mass) or of different nature (e.g., continuous and count variables), geometry of the trait space will be of questionable interpretability. Standardisation of traits by their mean or standard deviation would make traits nominally dimensionless, but it is generally an open question whether this procedure ensures interpretability of vectors and angles, which needs to be assessed in individual analyses.

The concentration of random angles around the right angle is just one of the potentially counterintuitive properties of high-dimensional spaces. Other superficially well-known concepts, such as volumes, Euclidean distances, and shapes of cubes and hyperspheres, also show peculiar behaviours in high-dimensional spaces (see Blum *et al.*, 2020; Rohlf, 2021). If biologists are to explore evolution in high-dimensional spaces, they ought to well familiarise themselves with properties of the quantities they work with. Even if the primary interest is in biology rather than geometry, basic geometric properties of the trait space should not be ignored.

5 Acknowledgements

The author would like to thank Yoel E. Stuart for kindly permitting partial reuse and redistribution of codes for the re-analysis in this paper. This work was partly supported by the Newton International Fellowships by the Royal Society (NIF\R1\180520) and the Overseas Research Fellowships by the Japan Society for the Promotion of Science (202160529). The author declares no conflict of interest.

Appendix A Comment on eigenanalysis

A.1 Singular value decomposition

De Lisle & Bolnick (2020) proposed to capture the magnitude of parallelism via analysis of eigenvalues of $\mathbf{ZZ}^T$. Here, it is argued that this procedure should be based on the eigenanalysis of $\mathbf{Z}^T\mathbf{Z}$ instead (or, more conveniently, singular value decomposition of $\mathbf{Z}$) by reviewing well-known results of matrix analysis. Readers familiar with singular value decomposition may simply skip this subsection. In the text, $\mathbf{Z}$ was defined as the $n \times p$ (lineage $\times$ trait) matrix, with the rows standardised to have the unit length, so that $\mathbf{C} = \mathbf{ZZ}^T$ is interpreted as an $n \times n$ inter-lineage correlation matrix. Let its spectral decomposition (or eigendecomposition) be

$$\mathbf{C} = \mathbf{ULU}^T, \tag{A.1}$$

where $\mathbf{U}$ is an $n \times n$ matrix of eigenvectors and $\mathbf{L} = \mathrm{diag}(l_i)$ is an $n \times n$ diagonal matrix of eigenvalues l_i. The symmetry of $\mathbf{C}$ ensures the orthonormality of $\mathbf{U}$: $\mathbf{U}^T\mathbf{U} = \mathbf{UU}^T = \mathbf{I}_n$ (or $\mathbf{U}^{-1} = \mathbf{U}^T$; e.g., Schott, 2016). The eigenvectors $\mathbf{U}$ represent major axes of variation in the (standardised) n-dimensional lineage space, each of whose coordinate axes represents trait values of a lineage, whereas the eigenvalues describe the variances along these axes. Note that this is not in the trait space, whose coordinate axes represent traits.

On the other hand, define the $p \times p$ inter-trait cross-product matrix $\mathbf{A} = \mathbf{Z}^T\mathbf{Z}$ and its spectral decomposition

$$\mathbf{A} = \mathbf{VKV}^T, \tag{A.2}$$

where $\mathbf{V}$ is a $p \times p$ matrix of eigenvectors and $\mathbf{K} = \mathrm{diag}(k_i)$ is a $p \times p$ diagonal matrix of eigenvalues k_i. The eigenvectors $\mathbf{V}$ represent major axes of variation in the p-dimensional trait space, whereas the eigenvalues are sums of squares (n times variances) along these axes. Right-multiplication with $\mathbf{V}$ projects $\mathbf{Z}$ onto the p-dimensional principal component (PC) space, so that $\mathbf{ZV}$ is an $n \times p$ matrix of PC scores for the n lineages. (We could equivalently consider $\mathbf{A}/n$ as the $p \times p$ inter-trait covariance matrix, centred at the origins of phenotypic change vectors rather than sample means.) Clearly, if we are interested in variation in the trait space as intended by De Lisle & Bolnick (2020), attention should be directed to the eigendecomposition of $\mathbf{A}$, which pertains to the trait space, rather than that of $\mathbf{C}$, which pertains to the lineage space. The eigenanalysis of $\mathbf{A}$ is also known in directional statistics as the analysis of principal directions (Mardia *et al.*, 1979).

In fact, it is to be seen that, from $\mathbf{CU} = \mathbf{ZZ}^T\mathbf{U} = \mathbf{UL}$ (eq. A.1), $\mathbf{A}(\mathbf{Z}^T\mathbf{U}) = \mathbf{Z}^T\mathbf{ZZ}^T\mathbf{U} = (\mathbf{Z}^T\mathbf{U})\mathbf{L}$, and hence $\mathbf{Z}^T\mathbf{U}$ is a matrix of eigenvectors for $\mathbf{A}$, with non-zero elements of $\mathbf{L}$ being common eigenvalues of $\mathbf{C}$ and $\mathbf{A}$ (see, e.g., Schott, 2016). In other words, for the eigenvalues of $\mathbf{C}$ and $\mathbf{A}$ (arranged in decreasing order),

$$\begin{cases} (l_1, \ldots, l_n, 0, \ldots, 0) = (k_1, \ldots, k_n, k_{n+1}, \ldots, k_p), & n < p. \\ (l_1, \ldots, l_n) = (k_1, \ldots, k_p), & n = p. \\ (l_1, \ldots, l_p, l_{p+1}, \ldots, l_n) = (k_1, \ldots, k_p, 0, \ldots, 0), & n > p. \end{cases} \tag{A.3}$$

This leads to a useful matrix factorisation known as singular value decomposition. For the $n \times p$ matrix $\mathbf{Z}$ of non-zero rank, we can find the decomposition

$$\mathbf{Z} = \mathbf{U}\mathbf{D}\mathbf{V}^T, \tag{A.4}$$

where $\mathbf{U}$ and $\mathbf{V}$ are $n \times n$ and $p \times p$, respectively, orthogonal matrices. $\mathbf{D}$ is an $n \times p$ matrix with the form:

$$\begin{cases} \begin{pmatrix} \boldsymbol{\Delta} & \mathbf{0} \end{pmatrix}, & n \leq p \\ \begin{pmatrix} \boldsymbol{\Delta} \\ \mathbf{0} \end{pmatrix}, & n \geq p \end{cases}$$

where $\boldsymbol{\Delta}$ is a diagonal matrix with the dimensionality $\min(n, p)$, and $\mathbf{0}$ denotes a matrix of 0's of an appropriate dimension (which disappears when $n = p$). The diagonal elements of $\boldsymbol{\Delta}$ are called singular values of $\mathbf{Z}$ (which can contain one or more 0's if and only if $\mathbf{Z}$ is rank-deficient). Observe

$$\mathbf{Z}\mathbf{Z}^T = \mathbf{U}\mathbf{D}\mathbf{V}^T\mathbf{V}\mathbf{D}^T\mathbf{U}^T = \mathbf{U}\mathbf{D}\mathbf{D}^T\mathbf{U}^T, \tag{A.5}$$

$$\mathbf{Z}^T\mathbf{Z} = \mathbf{V}\mathbf{D}^T\mathbf{U}^T\mathbf{U}\mathbf{D}\mathbf{V}^T = \mathbf{V}\mathbf{D}^T\mathbf{D}\mathbf{V}^T, \tag{A.6}$$

so that the decompositions of equations A.1 and A.2 can be constructed with the same $\mathbf{U}$ and $\mathbf{V}$, by letting $\mathbf{D}\mathbf{D}^T = \mathbf{L}$ and $\mathbf{D}^T\mathbf{D} = \mathbf{K}$. In other words, the singular values are square roots of the non-zero eigenvalues considered above. This result is often utilised in practical calculation of PCs (Jolliffe, 2002). It also clarifies how eigenvectors of $\mathbf{C}$ can be 'related back' to the trait space (De Lisle & Bolnick, 2020: equation 7 and figure 2):

$$\mathbf{Z}^T\mathbf{U}^{T^{-1}} = \mathbf{Z}^T\mathbf{U} = \mathbf{V}\mathbf{D}^T. \tag{A.7}$$

This $p \times n$ matrix equals the eigenvectors $\mathbf{V}$ of $\mathbf{A}$ scaled to have the lengths equal to square root of the corresponding eigenvalues. Although eigenvectors of $\mathbf{C}$ convey information regarding how the lineages are loaded on PCs, the same information can be obtained from PC scores with respect to $\mathbf{A}$, namely, $\mathbf{Z}\mathbf{V} = \mathbf{U}\mathbf{D}$. It may be worth noting that De Lisle & Bolnick's (2020) original suggestion is essentially to conduct principal coordinate analysis with $\mathbf{C}$ as the inter-lineage similarity matrix, which yields the same coordinates $\mathbf{U}\mathbf{D}$ in this case. In any case, all the relevant information can be extracted from singular value decomposition of $\mathbf{Z}$.

A.2 Eigenvalue dispersion and squared correlations

The equivalence between eigenvalue dispersion and sum of squared correlations mentioned in equation 15 can be confirmed as follows. Let r_{ij} be the (i, j)-th elements of the correlation matrix $\mathbf{C}$ (the diagonal elements are unity; $r_{ii} = 1$ for all i), and $\mathrm{tr}(\cdot)$ denote the matrix trace operator (sum of diagonal elements). It is easily confirmed that $\mathrm{tr}(\mathbf{S}\mathbf{T}) = \mathrm{tr}(\mathbf{T}\mathbf{S})$ holds for any arbitrary matrices $\mathbf{S}$ and $\mathbf{T}$ where the products can be defined (e.g., Schott, 2016), and hence $n = \sum_{i=1}^{n} r_{ii} = \mathrm{tr}(\mathbf{C}) = \mathrm{tr}(\mathbf{U}\mathbf{L}\mathbf{U}^T) = \mathrm{tr}(\mathbf{L}\mathbf{U}^T\mathbf{U}) = \mathrm{tr}(\mathbf{L}) = \sum_{i=1}^{n} l_i = n\bar{l}$, where $\bar{l}$ is the average

of eigenvalues (which equals 1). Then,

$$\begin{aligned} 2\sum_{i<j}^{n} r_{ij}^2 &= \sum_{i\neq j}^{n} r_{ij}^2 = \sum_{i,j=1}^{n} r_{ij}^2 - \sum_{i=1}^{n} r_{ii}^2 = \mathrm{tr}(\mathbf{C}^2) - n \\ &= \mathrm{tr}(\mathbf{ULU}^T\mathbf{ULU}^T) - n \\ &= \mathrm{tr}(\mathbf{L}^2) - n = \sum_{i=1}^{n} l_i^2 - 2\sum_{i=1}^{n} l_i + n \\ &= \sum_{i=1}^{n} \left(l_i - \bar{l}\right)^2. \end{aligned} \tag{A.8}$$

Equation A.8 holds for any correlation matrix in general (e.g., Gleason & Staelin, 1975; Durand & Le Roux, 2017; Watanabe, 2021). Also, by equation A.3, $\sum_{i=1}^{n} l_i^m = \sum_{i=1}^{p} k_i^m$ holds for any integer m. Therefore,

$$\begin{aligned} \sum_{i=1}^{p} \left(k_i - \bar{k}\right)^2 &= \sum_{i=1}^{p} k_i^2 - \frac{1}{p}\left(\sum_{i=1}^{p} k_i\right)^2 \\ &= \sum_{i=1}^{n} l_i^2 - \frac{n^2}{p} \\ &= \sum_{i=1}^{n} \left(l_i - \bar{l}\right)^2 + n - \frac{n^2}{p}. \end{aligned} \tag{A.9}$$

Equations A.8 and A.9 together indicate that the sum of squares of the eigenvalues of $\mathbf{C}$ and $\mathbf{A}$ around the respective averages are simple linear functions of each other and of sum of squared correlations in $\mathbf{C}$ (eq. 15).

A.3 Dimensionality of parallelism?

In proposing their eigenanalysis framework, De Lisle & Bolnick (2020) declared two objectives: quantifying the magnitude of parallelism; and determining the dimensionality of parallelism. The present study primarily concerns the first point, partly because working on the second point requires much more conceptual and methodological elaboration than can be presented here. Nevertheless, it seems pertinent to make a cautionary note; in short, it is rather obscure what quantity their original procedure aims to evaluate, beyond testing the null hypothesis of no concentration of phenotypic change vectors.

Problematically, De Lisle & Bolnick (2020) did not formally state their hypotheses for identifying the dimensionality of parallelism or how they can be tested, be it with statistical null-hypothesis testing or any other framework. They mentioned comparing observed eigenvalues with Monte Carlo distributions, but were not explicit as to which eigenvalue is compared in what criterion. Nevertheless, their writing seems to imply that all observed eigenvalues are compared with the null distributions of eigenvalues generated with Monte Carlo simulations, and that the eigenvalues larger than the null distributions are considered 'significant'. Their allusion to the number of 'significant' eigenvalues is reminiscent of the traditional component retention framework in PCA, where a small number of leading PCs (those corresponding to

large eigenvalues) are retained while the others are discarded (e.g., Jolliffe, 2002; Peres-Neto *et al.*, 2005; Schott, 2006; Dray, 2008; Vieira, 2012; Björklund, 2019). The premise there is that trailing PCs tend to be dominated by noise—most typically, measurement error—and hence are unlikely to be meaningfully interpreted. For this purpose, it is insensible to compare all observed eigenvalues with Monte Carlo distributions simultaneously, for the following two reasons. First, due to the scaling of $\mathbf{Z}$, the eigenvalues of $\mathbf{A}$ (or $\mathbf{C}$) are constrained to sum to n (see above). This means that, if there is any detectably large eigenvalue, other ones are necessarily small, even if the PCs corresponding to the latter may still convey informative signals. Hence, potentially meaningful information can be lost if eigenvalues smaller than the simultaneous null distributions are discarded altogether. Second, it is conceivable that, say, the second and third eigenvalues are smaller and larger, respectively, than their respective critical points. Although such a situation may arguably be of rare occurrence in practice, it is not obvious which components are to be retained or discarded. The simultaneous comparison is valid for testing the null hypothesis $\mathbf{\Gamma} = \mathbf{I}_n$, provided that the appropriate null distributions are used (see text), but is unlikely to serve for determining the number of 'significant' eigenvalues.

If one is interested in determining the number of components that exceed the magnitude of noise and hence are interpretable on their own, then a potential option is to use one of the component retention methods (see papers cited above), perhaps with appropriate modifications. However, most of these methods proceed with stepwise null-hypothesis testing—test of the equality of adjacent population eigenvalues—at the risk of committing type I or type II error at each step. Therefore, the resultant number of significant eigenvalues would need to be viewed with much caution. Inaccuracy in the number of significant eigenvalues is not a serious concern in ordinary applications of PCA since the primary goal is usually to obtain a convenient low-dimensional approximation, but it will be a crucial issue if the number of components itself is the target of inquiry.

Determination of matrix dimensionality is also a matter of interest in quantitative genetics, where a lack of full dimensionality in an additive genetic covariance matrix is regarded as evidence for genetic constraint (e.g., Kirkpatrick, 2009; Hine *et al.*, 2014). Adopting one of the techniques there (Hine & Blows, 2006; Pavlicev *et al.*, 2009b) might potentially be another option for determining the dimensionality of parallelism. Although superficially similar to the previous problem of component retention, they test different statistical hypotheses and typically require an estimate of error.

In any case, determination of the dimensionality of parallelism requires more conceptual clarification and methodological rigour. It is beyond the scope here to elaborate on potential adequacy of all the different methods mentioned above. Apart from the determination of dimensionality, it would be worth remembered that reduced-rank inferences are prone to induce inaccuracy in many biological settings (see Meyer & Kirkpatrick, 2008; Uyeda *et al.*, 2015; Adams & Collyer, 2018). Thus, for many purposes, treating high-dimensional data as they are seems a favourable option in general.

Appendix B Dataset for example analysis

In the text, Stuart *et al.*'s (2017) dataset of lake–stream divergence in the threespine stickleback (*Gasterosteus aculeatus*) is re-analysed. The dataset analysed here consists of 80 nominal morphological traits (see text). The 4 meristic traits originally included in the dataset were excluded from the present analysis because mixing different types of traits might compromise interpretability (although results were qualitatively similar even when these were involved). Raw data and codes were retrieved from http://web.corral.tacc.utexas.edu/Stuart_2017_NatureEE_Data_Code/ on 17th August 2021. Individuals with any missing data entries were excluded from analysis. As a result, the vectors for 13 watersheds, out of the original 16, were retained; the excluded localities were Boot, Pye, and Village Bay, which lacked complete observations in at least one of the ancestor or descendant population. The variables were standardised by the pooled standard deviations across all localities, rather than by the standard deviations for each watershed as was originally done in Stuart *et al.* (2017). These procedures are primarily for ease of the demonstrative analyses, and not to be advocated for methodological adequacy. Codes to reproduce the re-analysis are provided as Supplemental Material.

Appendix C Exploratory analysis

This section briefly discusses exploratory and visualisation methods for phenotypic change vectors. It is straightforward to treat phenotypic change vectors—rows of $\mathbf{X}$ or $\mathbf{Z}$ defined in the text—as 'observations', as has been done for the analysis of allometric space (Klingenberg & Froese, 1991; Gerber *et al.*, 2008; Gerber & Hopkins, 2011). The space of phenotypic trajectories could be called trajectory space. When the vectors are normalised to have unit length, the resultant space is on the unit hypersphere in the p-dimensional space. Visualisation can be achieved via dimension reduction or ordination methods. For example, eigenanalysis of the $p \times p$ (inner) cross-product matrix $\mathbf{A}$ or equivalently covariance matrix $\mathbf{A}/n$ can be used to visualise variation in phenotypic change vectors via PCA. As defined in the text, these are centred at the origins of each phenotypic change vector. Alternatively, the columns of $\mathbf{Z}$ could be mean-centred, or $\mathbf{A}$ could be scaled as a correlation matrix, each giving a different interpretation. It would be useful to visualise estimates of sampling errors or confidence regions for trajectories by plotting these quantities at the same time (Klingenberg & Froese, 1991; Fig. 3E). One caveat is that typical 2- or 3-dimensional plots usually give distorted visualisations of the space, especially when the vectors are widely spread across the hypersphere. For this reason, it is always advisable to supplement this visualisation with inspection of the angles/distances (e.g., Fig. 3D). Specialised methods for directional data (reviewed in Pewsey & García-Portugués, 2021) may perhaps help mitigate such problems.

Biplots provide useful, simultaneous visualisations of relationships between observations and variables (see, e.g., Jolliffe, 2002; Greenacre, 2012; Gower *et al.*, 2015). This technique is closely related to singular value decomposition mentioned in Appendix A. Let q denote the rank of the data matrix $\mathbf{Z}$. Typically, we consider a factorisation of the form

$$\mathbf{Z} = \mathbf{U}\mathbf{D}\mathbf{V}^T = \mathbf{U}_q\mathbf{\Delta}\mathbf{V}_q^T = \mathbf{U}_q\mathbf{\Delta}^\alpha\mathbf{\Delta}^{1-\alpha}\mathbf{V}_q^T, \tag{C.1}$$

where $\mathbf{U}$, $\mathbf{D}$, $\mathbf{V}$ and are as in equation A.4, $\mathbf{U}_q$ and $\mathbf{V}_q$ are the matrices containing the q columns of $\mathbf{U}$ and $\mathbf{V}$, respectively, corresponding to non-zero entries of $\mathbf{D}$ or $\mathbf{\Delta}$, and $\mathbf{\Delta}^\alpha$ and $\mathbf{\Delta}^{1-\alpha}$ are diagonal matrices whose diagonal elements are αth and $(1-\alpha)$th powers, respectively, of those of $\mathbf{\Delta}$, with the scaling parameter α usually taken within $[0, 1]$. If we denote $\mathbf{G} = \mathbf{U}_q\mathbf{\Delta}^\alpha$ and $\mathbf{H} = \mathbf{V}_q\mathbf{\Delta}^{1-\alpha}$, $\mathbf{G}$ and $\mathbf{H}$ convey information pertaining to the lineages and traits, respectively. The columns of $\mathbf{G}$ and $\mathbf{H}$ corresponding to the largest singular values provide the best lower-dimensional approximation of the original data, enabling a useful visualisation. When $\alpha = 0$, the rows of $\mathbf{G}$ do not preserve Euclidean distances between lineages (but are instead in the Mahalanobis metric), whereas those of $\mathbf{H}$ tend to preserve the covariance structure between traits (cosines are correlations). When $\alpha = 1$, the rows of $\mathbf{G}$ (ordinary PC scores) preserve Euclidean metric relationships between lineages, but those of $\mathbf{H}$ (ordinary PC coefficients) do not exactly preserve the covariance structure (for details see Jolliffe, 2002). This technique can be applied to the analysis of phenotypic change vectors (Fig. 3C). It should be noted that traits like the Procrustes shape coordinates can be visualised in this way but cannot be interpreted individually apart from one another (e.g., Klingenberg, 2013). The procedure described here belongs to the classical PCA biplot, and there are many variants corresponding to other multivariate analytic techniques, which may turn useful in certain situations (see Gower *et al.*, 2015).

Klingenberg & Froese (1991) proposed to detect patterns in a set of vectors by applying the concept of principal points. Principal points are a given number of points in a high-dimensional space that collectively minimise expected squared Euclidean distances from (the nearest one of) them in a population (Flury, 1990). In practice, principal points are typically estimated with k-means clustering, which seeks k such groups of observations that minimise within-group sum of squares in a sample. Depending on assumptions on the underlying distribution, clustering on constrained subspaces may yield better estimates of principal points (Flury, 1993). Although it is beyond the scope here to present an extensive technical overview of various clustering methods (see e.g., Steinley, 2006), this framework seems to have broad applications in the analysis of phenotypic change vectors, such as detecting optimal clustering or obtaining representative vectors from a large number of vectors.

In the present re-analysis, PCA of $\mathbf{Z}$ (without mean-centring of columns) was used for visualisation of the trajectory space (Fig. 3C–F). PCA biplot with $\alpha = 1$ is shown in Figure 3C. Sampling error in the trajectories were assessed with nonparametric bootstrapping with 5000 replicates, in which individuals were randomly drawn with replacement to match each within-locality sample size. Only 1000 replicates are plotted in Figure 3E for visual clarity. As the vectors are lying on the surface of a hypersphere, the 2-dimensional plot is inevitably distorted. The 95% confidence regions based on the first 2 PC scores of full 5000 replicates are plotted as ellipses for visual aid, although the assumption of multivariate normality may not be particularly appropriate here. Grouping of the vectors were explored with k-means clustering of the entire data $\mathbf{Z}$, via Hartigan and Wong's (1979) algorithm implemented in the function '`kmeans`' of `R` environment version 3.5.3 (R Core Team, 2019). Clustering performance was assessed by within-group sum of squares. In order to avoid falling into local minima, clustering was conducted repeatedly, with all possible combinations of observations used as initial group means, following Klingenberg & Froese (1991). For each k (varying from 2 to $n-1$), the clustering with smallest

within-group sum of squares was retained. Visual inspection suggested that within-group sum of squares continued to drop up to $k = 5$ but did not decrease substantially past this point. Hence, this clustering result is presented in Figure 3F as an example of exploratory analysis, although the visual inspection is not particularly adequate for determining optimal k (Steinley, 2006).

References

Adams, D. C., **Collyer, M. L.**, **2007**. Analysis of character divergence along environmental gradients and other covariates. *Evolution*, **61**: 510–515. doi:10.1111/j.1558-5646.2007.00063.x.

Adams, D. C., **Collyer, M. L.**, **2009**. A general framework for the analysis of phenotypic trajectories in evolutionary studies. *Evolution*, **63**: 1143–1154. doi:10.1111/j.1558-5646.2009.00649.x.

Adams, D. C., **Collyer, M. L.**, **2018**. Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. *Systematic Biology*, **67**: 14–31. doi:10.1093/sysbio/syx055.

Adams, D. C., **Collyer, M. L.**, **2019**. Phylogenetic comparative methods and the evolution of multivariate phenotypes. *Annual Review of Ecology, Evolution, and Systematics*, **50**: 405–425. doi:10.1146/annurev-ecolsys-110218-024555.

Agrawal, A. A., **2017**. Toward a predictive framework for convergent evolution: integrating natural history, genetic mechanisms, and consequences for the diversity of life. *American Naturalist*, **190**: S1–S12. doi:10.1086/692111.

Anderson, M. J., **2001**. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, **26**: 32–46. doi:10.1111/j.1442-9993.2001.01070.pp.x.

Anderson, T. W., **1963**. Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics*, **34**: 122–148. doi:10.1214/aoms/1177704248.

Anderson, T. W., **2003**. *An Introduction to Multivariate Statistical Analysis*. 3rd edition. John Wiley & Sons, Hoboken, New Jersey.

Arbuckle, K., **Bennett, C. M.**, **Speed, M. P.**, **2014**. A simple measure of the strength of convergent evolution. *Methods in Ecology and Evolution*, **5**: 685–693. doi:10.1111/2041-210X.12195.

Baliga, V. B., **Mehta, R. S.**, **2018**. Phylo-allometric analyses showcase the interplay between life-history patterns and phenotypic convergence in cleaner wrasses. *American Naturalist*, **191**: E129–E143. doi:10.1086/697047.

Björklund, M., **2019**. Be careful with your principal components. *Evolution*, **73**: 2151–2158. doi:10.1111/evo.13835.

Blount, Z. D., **Lenski, R. E.**, **Losos, J. B.**, **2018**. Contingency and determinism in evolution: replaying life's tape. *Science*, **362**: eaam5979. doi:10.1126/science.aam5979.

Blum, A., **Hopcroft, J.**, **Kannan, R.**, **2020**. *Foundations of Data Science*. Cambridge University Press, Cambridge, UK. doi:10.1017/9781108755528.

Bolnick, D. I., **Barrett, R. D. H.**, **Oke, K. B.**, **Rennison, D. J.**, **Stuart, Y. E.**, **2018**. (Non)parallel evolution. *Annual Review of Ecology, Evolution, and Systematics*, **49**: 303–330. doi:10.1146/annurev-ecolsys-110617-062240.

Cai, T., **Fan, J.**, **Jiang, T.**, **2013**. Distributions of angles in random packing on spheres. *Journal of Machine Learning Research*, **14**: 1837–1864. URL: https://www.jmlr.org/papers/volume14/cai13a/cai13a.pdf.

Cheverud, J. M., **Rutledge, J. J.**, **Atchley, W. R.**, **1983**. Quantitative genetics of development: genetic correlations among age-specific trait values and the evolution of ontogeny. *Evolution*, **37**: 895–905. doi:10.1111/j.1558-5646.1983.tb05619.x.

Collyer, M. L., **Adams, D. C.**, **2007**. Analysis of two-state multivariate phenotypic change in ecological studies. *Ecology*, **88**: 683–692. doi:10.1890/06-0727.

Collyer, M. L., **Adams, D. C.**, **2013**. Phenotypic trajectory analysis: comparison of shape change patterns in evolution and ecology. *Hystrix*, **24**: 75–83. doi:10.4404/hystrix-24.1-6298.

De Lisle, S. P., **Bolnick, D. I.**, **2020**. A multivariate view of parallel evolution. *Evolution*, **74**: 1466–1481. doi:10.1111/evo.14035.

Dray, S., **2008**. On the number of principal components: a test of dimensionality based on measurements of similarity between matrices. *Computational Statistics and Data Analysis*, **52**: 2228–2237. doi:10.1016/j.csda.2007.07.015.

Durand, J.-L., **Le Roux, B.**, **2017**. Linkage index of variables and its relationship with variance of eigenvalues in PCA and MCA. *Statistica Applicata – Italian Journal of Applied Statistics*, **29**: 123–135. doi:10.26398/IJAS.0029-006.

Feiner, N., **Jackson, I. S. C.**, **Van der Cruyssen, E.**, **Uller, T.**, **2021**. A highly conserved ontogenetic limb allometry and its evolutionary significance in the adaptive radiation of *Anolis* lizards. *Proceedings of the Royal Society B: Biological Sciences*, **288**: 20210226. doi:10.1098/rspb.2021.0226.

Fisher, R. A., **1915**. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, **10**: 507–521. doi:10.1093/biomet/10.4.507.

Fisher, R. A., **1925**. Applications of "Student's" distribution. *Metron*, **5**: 90–104.

Flury, B. A., **1990**. Principal points. *Biometrika*, **77**: 33–41. doi:10.1093/biomet/77.1.33.

Flury, B. D., **1993**. Estimation of principal points. *Applied Statistics*, **42**: 139–151. doi:10.2307/2347416.

Gerber, S., **Eble, G. J.**, **Neige, P.**, **2008**. Allometric space and allometric disparity: a developmental perspective in the macroevolutionary analysis of morphological disparity. *Evolution*, **62**: 1450–1457. doi:10.1111/j.1558-5646.2008.00370.x.

Gerber, S., **Hopkins, M. J.**, **2011**. Mosaic heterochrony and evolutionary modularity: the trilobite genus *Zacanthopsis* as a case study. *Evolution*, **65**: 3241–3252. doi:10.1111/j.1558-5646.2011.01363.x.

Gerber, S., **Neige, P.**, **Eble, G. J.**, **2007**. Combining ontogenetic and evolutionary scales of morphological disparity: a study of early Jurassic ammonites. *Evolution and Development*, **9**: 472–482. doi:10.1111/j.1525-142X.2007.00185.x.

Gleason, T. C., **Staelin, R.**, **1975**. A proposal for handling missing data. *Psychometrika*, **40**: 229–252. doi:10.1007/BF02291569.

Gower, J. C., **Le Roux, N. J.**, **Gardner-Lubbe, S.**, **2015**. Biplots: quantitative data. *Wiley Interdisciplinary Reviews Computational Statistics*, **7**: 42–62. doi:10.1002/wics.1338.

Greenacre, M. J., **2012**. Biplots: the joy of singular value decomposition. *Wiley Interdisciplinary Reviews Computational Statistics*, **4**: 399–406. doi:10.1002/wics.1200.

Haber, A., **2011**. A comparative analysis of integration indices. *Evolutionary Biology*, **38**: 476–488. doi:10.1007/s11692-011-9137-4.

Haines, G. E., **Stuart, Y. E.**, **Hanson, D.**, **Tasneem, T.**, **Bolnick, D. I.**, **Larsson, H. C. E.**, **Hendry, A. P.**, **2020**. Adding the third dimension to studies of parallel evolution of morphology and function: an exploration based on parapatric lake-stream stickleback. *Ecology and Evolution*, **10**: 13297–13311. doi:10.1002/ece3.6929.

Hartigan, J. A., **Wong, M. A.**, **1979**. A *k*-means clustering algorithm. *Journal of the Royal Statistical Society, Series C: Applied Statistics*, **28**: 100–108. doi:10.2307/2346830.

Hine, E., **Blows, M. W.**, **2006**. Determining the effective dimensionality of the genetic variance–covariance matrix. *Genetics*, **173**: 1135–1144. doi:10.1534/genetics.105.054627.

Hine, E., **McGuigan, K.**, **Blows, M. W.**, **2014**. Evolutionary constraints in high-dimensional trait sets. *American Naturalist*, **184**: 119–131. doi:10.1086/676504.

Hotelling, H., **1953**. New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society, Series B: Methodological*, **15**: 193–232. doi:10.1111/j.2517-6161.1953.tb00135.x.

James, M. E., **Wilkinson, M. J.**, **Bernal, D. M.**, **Liu, H.**, **North, H. L.**, **Engelstädter, J.**, **Ortiz-Barrientos, D.**, **2021**. Phenotypic and genotypic parallel evolution in parapatric ecotypes of *Senecio*. *bioRxiv*: 2020.02.05.936450. doi:10.1101/2020.02.05.936450.

Jolicoeur, P., **1963**a. The degree of generality of robustness in *Martes americana*. *Growth*, **27**: 1–27.

Jolicoeur, P., **1963**b. The multivariate generalization of the allometry equation. *Biometrics*, **19**: 497–499. doi:10.2307/2527939.

Jolliffe, I. T., **2002**. *Principal Component Analysis*. 2nd edition. Springer, New York, New York. doi:10.1007/b98835.

Kaeuffer, R., **Peichel, C. L.**, **Bolnick, D. I.**, **Hendry, A. P.**, **2012**. Parallel and nonparallel aspects of ecological, phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. *Evolution*, **66**: 402–418. doi:10.1111/j.1558-5646.2011.01440.x.

Kirkpatrick, M., **2009**. Patterns of quantitative genetic variation in multiple dimensions. *Genetica*, **136**: 271–284. doi:10.1007/s10709-008-9302-6.

Klingenberg, C. P., **1996**. Multivariate allometry. *In:* Marcus, L. F. (ed.), *Advances in Morphometrics*. Plenum Press, New York, New York, pp. 23–49. doi:10.1007/978-1-4757-9083-2_3.

Klingenberg, C. P., **1998**. Heterochrony and allometry: the analysis of evolutionary change in ontogeny. *Biological Reviews*, **73**: 79–123. doi:10.1017/S000632319800512X.

Klingenberg, C. P., **2013**. Visualizations in geometric morphometrics: how to read and how to make graphs showing shape changes. *Hystrix*, **24**: 15–24. doi:10.4404/hystrix-24.1-7691.

Klingenberg, C. P., **2016**. Size, shape, and form: concepts of allometry in geometric morphometrics. *Development Genes and Evolution*, **226**: 113–137. doi:10.1007/s00427-016-0539-2.

Klingenberg, C. P., **Froese, R.**, **1991**. A multivariate comparison of allometric growth patterns. *Systematic Zoology*, **40**: 410–419. doi:10.1093/sysbio/40.4.410.

Klingenberg, C. P., **Marugán-Lobón, J.**, **2013**. Evolutionary covariation in geometric morphometric data: analyzing integration, modularity, and allometry in a phylogenetic context. *Systematic Biology*, **62**: 591–610. doi:10.1093/sysbio/syt025.

Langerhans, R. B., **2018**. Predictability and parallelism of multitrait adaptation. *Journal of Heredity*, **109**: 59–70. doi:10.1093/jhered/esx043.

Ley, C., **Verdebout, T.**, **2017**. *Modern Directioanl Statistics*. CRC Press, Boca Raton, Florida.

Li, S., **2011**. Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics and Statistics*, **4**: 66–70. doi:10.3923/ajms.2011.66.70.

Manly, B. F. J., **Navarro Alberto, J. A. J.**, **2020**. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 4th edition. Chapman and Hall, Boca Raton, Florida.

Mardia, K. V., **Jupp, P. E.**, **1999**. *Directional Statistics*. John Wiley & Sons, Chichester, UK. doi:10.1002/9780470316979.

Mardia, K. V., **Kent, J. T.**, **Bibby, J. M.**, **1979**. *Multivariate Analysis*. Academic Press, London, UK.

Meyer, K., **Kirkpatrick, M.**, **2008**. Perils of parsimony: properties of reduced-rank estimates of genetic covariance matrices. *Genetics*, **180**: 1153–1166. doi:10.1534/genetics.108.090159.

Mitteroecker, P., **Gunz, P.**, **Bookstein, F. L.**, **2005**. Heterochrony and geometric morphometrics: a comparison of cranial growth in *Pan paniscus* versus *Pan troglodytes*. *Evolution and Development*, **7**: 244–258. doi:10.1111/j.1525-142X.2005.05027.x.

Muirhead, R. J., **1982**. *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, Hoboken, New Jersey. doi:10.1002/9780470316559.

Oke, K. B., **Rolshausen, G.**, **LeBlond, C.**, **Hendry, A. P.**, **2017**. How parallel is parallel evolution? A comparative analysis in fishes. *American Naturalist*, **190**: 1–16. doi:10.1086/691989.

Olkin, I., **Finn, J.**, **1990**. Testing correlated correlations. *Psychological Bulletin*, **108**: 330–333. doi:10.1037/0033-2909.108.2.330.

Olkin, I., **Finn, J. D.**, **1995**. Correlations redux. *Psychological Bulletin*, **118**: 155–164. doi:10.1037/0033-2909.118.1.155.

Olkin, I., **Siotani, M.**, **1976**. Asymptotic distribution of functions of a correlation matrix. *In:* Editorial Committee for Publication of Essays in Probability and Statistics (ed.), *Essays in Probability and Statistics in Honor of Professor Junjiro Ogawa*. Shinko Tsusho, Tokyo, Japan, pp. 235–251.

Owens, G. L., **Veen, T.**, **Moxley, D. R.**, **Arias-Rodriguez, L.**, **Tobler, M.**, **Renninson, D. J.**, **2021**. Parallel shifts of visual sensitivity and body colouration in replicate populations of extremophile fish. *bioRxiv*: 2021.06.16.448734. doi:10.1101/2021.06.16.448734.

Pavlicev, M., **Cheverud, J. M.**, **Wagner, G. P.**, **2009**a. Measuring morphological integration using eigenvalue variance. *Evolutionary Biology*, **36**: 157–170. doi:10.1007/s11692-008-9042-7.

Pavlicev, M., **Wagner, G. P.**, **Cheverud, J. M.**, **2009**b. Measuring evolutionary constraints through the dimensionality of the phenotype: adjusted bootstrap method to estimate rank of phenotypic covariance matrix. *Evolutionary Biology*, **36**: 339–353. doi:10.1007/s11692-009-9066-7.

Peres-Neto, P. R., **Jackson, D. A.**, **Somers, K. M.**, **2005**. How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics and Data Analysis*, **49**: 974–997. doi:10.1016/j.csda.2004.06.015.

Pérez-Ben, C. M., **Báez, A. M.**, **Schoch, R. R.**, **2020**. Morphological evolution of the skull roof in temnospondyl amphibians mirrors conservative ontogenetic patterns. *Zoological Journal of the Linnean Society*, **188**: 163–179. doi:10.1093/zoolinnean/zlz068.

Pewsey, A., **García-Portugués, E.**, **2021**. Recent advances in directional statistics. *TEST*, **30**: 1–58. doi:10.1007/s11749-021-00759-x.

R Core Team, **2019**. R: a language and environment for statistical computing, version 3.5.3. URL: `https://www.R-project.org/`.

Radersma, R., **Noble, D. W. A.**, **Uller, T.**, **2020**. Plasticity leaves a phenotypic signature during local adaptation. *Evolution Letters*, **4**: 360–370. doi:10.1002/evl3.185.

Raeymaekers, J. A. M., **Chaturvedi, A.**, **Hablützel, P. I.**, **Verdonick, I.**, **Hellermans, B.**, **Maes, G. E.**, **De Meester, L.**, **Volckaert, F. A. M.**, **2017**. Adaptive and non-adaptive divergence in a common landscape. *Nature Communications*, **8**: 267. doi:10.1038/s41467-017-00256-6.

Ram, Y., **Hadany, L.**, **2015**. The probability of improvement in Fisher's geometric model: a probabilistic approach. *Theoretical Population Biology*, **99**: 1–6. doi:10.1016/j.tpb.2014.10.004.

Rice, S. H., **1990**. A geometric model for the evolution of development. *Journal of Theoretical Biology*, **143**: 319–342. doi:10.1016/S0022-5193(05)80033-5.

Rincon-Sandoval, M., **Duarte-Ribeiro, E.**, **Davis, A. M.**, **Santaquiteria, A.**, **Hughes, L. C.**, **Baldwin, C. C.**, **Soto-Torres, L.**, **Acero P., A.**, **Walker, H. J., Jr.**, **Carpenter, K. E.**, **Sheaves, M.**, **Orti, G.**, **Arcila, D.**, **Betancur-R., R.**, **2020**. Evolutionary determinism and convergence associated with water-column transitions in marine fishes. *Proceedings of the National Academy of Sciences of the United States of America*, **117**: 33396–33403. doi:10.1073/pnas.2006511117.

Rohlf, F. J., **2021**. Why clusters and other patterns can seem to be found in analyses of high-dimensional data. *Evolutionary Biology*, **48**: 1–16. doi:10.1007/s11692-020-09518-6.

Rohner, P. T., **2020**. Evolution of multivariate wing allometry in schizophoran flies (Diptera: Schizophora). *Journal of Evolutionary Biology*, **33**: 831–841. doi:10.1111/jeb.13613.

Schott, J. R., **2005**. Testing for complete independence in high dimensions. *Biometrika*, **92**: 951–956. doi:10.1093/biomet/92.4.951.

Schott, J. R., **2006**. A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix. *Journal of Multivariate Analysis*, **97**: 827–843. doi:10.1016/j.jmva.2005.05.003.

Schott, J. R., **2016**. *Matrix Analysis for Statistics*. 3rd edition. John Wiley & Sons, Hoboken, New Jersey.

Shea, B. T., **1985**. Bivariate and multivariate growth allometry: statistical and biological considerations. *Journal of Zoology, Series A*, **206**: 367–390. doi:10.1111/j.1469-7998.1985.tb05665.x.

Sheets, H. D., **Zelditch, M. L.**, **2013**. Studying ontogenetic trajectories using resampling methods and landmark data. *Hystrix*, **24**: 67–73. doi:10.4404/hystrix-24.1-6332.

Solignac, M., **Cariou, M.-L.**, **Wimitzky, M.**, **1990**. Variability, specificity and evolution of growth gradients in the species complex *Jaera albifrons* (Isopoda, Asellota). *Crustaceana*, **59**: 121–145. URL: https://www.jstor.org/stable/20104588.

Speed, M. P., **Arbuckle, K.**, **2017**. Quantification provides a conceptual basis for convergent evolution. *Biological Reviews*, **92**: 815–829. doi:10.1111/brv.12257.

Stayton, C. T., **2006**. Testing hypotheses of convergence with multivariate data: morphological and functional convergence among herbivorous lizards. *Evolution*, **60**: 824–841. doi:10.1111/j.0014-3820.2006.tb01160.x.

Stayton, C. T., **2008**. Is convergence surprising? An examination of the frequency of convergence in simulated datasets. *Journal of Theoretical Biology*, **252**: 1–14. doi:10.1016/j.jtbi.2008.01.008.

Stayton, C. T., **2015**. The definition, recognition, and interpretation of convergent evolution, and two new measures for quantifying and assessing the significance of convergence. *Evolution*, **69**: 2140–2153. doi:10.1111/evo.12729.

Steinley, D., **2006**. K-means clustering: a half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, **59**: 1–34. doi:10.1348/000711005X48266.

Strelin, M. M., **Benitez-Vieyra, S.**, **Fornoni, J.**, **Klingenberg, C. P.**, **Cocucci, A.**, **2018**. The evolution of floral ontogenetic allometry in the Andean genus *Caiophora* (Loasaceae, subfam. Loasoideae). *Evolution and Development*, **20**: 29–39. doi:10.1111/ede.12246.

Stuart, A., **Ord, J. K.**, **1994**. *Kendall's Advanced Theory of Statistics*, volume 1. 6th edition. Hodder Education, London, UK. [Reprinted in 2004 by John Wiley & Sons, Chichester, UK].

Stuart, Y. E., **Veen, T.**, **Weber, J. N.**, **Hanson, D.**, **Ravinet, M.**, **Lohman, B. K.**, **Thompson, C. J.**, **Tasneem, T.**, **Doggett, A.**, **Izen, R.**, **Ahmed, N.**, **Barrett, R. D. H.**, **Hendryr, A. P.**, **Peichel, C. L.**, **Bolnick, D. I.**, **2017**. Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nature Ecology and Evolution*, **1**: 0158. doi:10.1038/s41559-017-0158.

Thompson, K. A., **Osmond, M. M.**, **Schluter, D.**, **2019**. Parallel genetic evolution and speciation from standing variation. *Evolution Letters*, **3**: 129–141. doi:10.1002/evl3.106.

Urošević, A., **Ljubisavljević, K.**, **Ivanović, A.**, **2013**. Patterns of cranial ontogeny in lacertid lizards: morphological and allometric disparity. *Journal of Evolutionary Biology*, **26**: 399–415. doi:10.1111/jeb.12059.

Uyeda, J. C., **Caetano, D. S.**, **Pennel, M. W.**, **2015**. Comparative analysis of principal components can be misleading. *Systematic Biology*, **64**: 677–689. doi:10.1093/sysbio/syv019.

Vieira, V. M. N. C. S., **2012**. Permutation tests to estimate significances on principal components analysis. *Computational Ecology and Software*, **2**: 103–123. URL: http://www.iaees.org/publications/journals/ces/articles/2012-2(2)/permutation-tests-to-estimate-significances.pdf.

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. doi:10.1007/BF00275224.

Watanabe, J., **2021**. Statistics of eigenvalue dispersion indices: quantifying the magnitude of phenotypic integration. *bioRxiv*: 2021.06.19.449119. doi:10.1101/2021.06.19.449119. [In press from *Evolution*].

Weber, A. A.-T., **Rajkov, J.**, **Smailus, K.**, **Egger, B.**, **Salzburger, W.**, **2021**. Diversification dynamics and (non-)parallel evolution along an ecological gradient in African cichlid fishes. *bioRxiv*: 2021.01.12.426414. doi:10.1101/2021.01.12.426414.

Wilson, L. A. B., **2013**. Allometric disparity in rodent evolution. *Ecology and Evolution*, **3**: 971–984. doi:10.1002/ece3.521.

Wilson, L. A. B., **2018**. The evolution of ontogenetic allometric trajectories in mammalian domestication. *Evolution*, **72**: 867–877. doi:10.1111/evo.13464.

Zelditch, M. L., **Calamari, Z. T.**, **Swiderski, D. L.**, **2016**. Disparate postnatal ontogenies do not add to the shape disparity of infants. *Evolutionary Biology*, **43**: 188–207. doi:10.1007/s11692-016-9370-y.

Zelditch, M. L., **Sheets, H. D.**, **Fink, W. L.**, **2003**. The ontogenetic dynamics of shape disparity. *Paleobiology*, **29**: 139–156. doi:10.1666/0094-8373(2003)029<0139:TODOSD>2.0.CO;2.

Zou, G. Y., **2007**. Toward using confidence intervals to compare correlations. *Psychological Methods*, **12**: 399–413. doi:10.1037/1082-989X.12.4.399.