Author for correspondence:
Junya Watanabe
e-mail: jw2098@cam.ac.uk

THE ROYAL SOCIETY
PUBLISHING

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, UK

JW, 0000-0002-9810-5286

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, properties of angles in multidimensional spaces have not been widely appreciated by biologists. 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 (i.e. uniform distribution on hypersphere), 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 can be connected to the test of no correlation or concentration of multiple vectors, for which simple test procedures are 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.

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 conceptualize and analyse patterns of phenotypic evolution in multidimensional trait spaces (e.g. [1–7]). However, increasing dimensionality sometimes poses challenges in interpreting and analysing quantities that superficially appear familiar. This review concerns technical 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. Here, the term parallel evolution is used in the geometric sense; parallelism between trajectories in a trait space between multiple ancestor–descendant pairs [7,8], 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, and the prevalence and extent of such parallelism are under active debate (e.g. [9–11]).

A variety of toolkit exists for analysing evolutionary or developmental trajectories in multidimensional spaces. One useful concept is the allometric space [12–16], where variation among multivariate allometric axes (typically principal component (PC) vectors; [17–21]) can be visualized and analysed in various

ways by treating empirical allometric axes as observations [22–29]. Another broadly employed tool is the phenotypic trajectory analysis [3,30–32], which primarily concerns quantification and statistical testing of inter-population differences in the magnitude, direction and shape of phenotypic trajectories.

The phenotypic trajectory (or phenotypic change vector) analysis has recently fuelled investigations into the parallel evolution [7,33]. There is a trend to quantitatively analyse patterns of evolutionary changes in putatively parallel lineages (e.g. [34–36]). The angles between phenotypic change vectors of different lineages play an especially pivotal role in empirical analyses of parallel evolution (e.g. [33,37–43]), because they are supposed to provide 'intuitive and mathematically formal' measures of (non)parallelism [37, p. 6].

Unfortunately, however, interpretation of angles in multidimensional spaces is not so straightforward. Consider, for example, the angle between randomly directed vectors in two- and three-dimensional spaces. It is convenient to fix one of them pointed at a 'pole' and to let the other be uniformly distributed on the unit circle/sphere (figure 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 two-dimensional space is uniform (figure 1a), whereas that for the three-dimensional space is peaked at the 'equator' because this region encompasses more area per latitude than 'polar' regions (figure 1b). This simple example demonstrates that distributions of random angles depend on the dimensionality, warning against extending our intuition into high-dimensional spaces. Regrettably, few recent analyses of evolutionary parallelism have 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 of 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 have no 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. [44–47]). The primary aim here is to disseminate well-known results with theoretical underpinnings. Recently, a framework for analysing multiple vectors simultaneously via eigenanalysis of a vector correlation matrix was proposed [48], but this framework lacked clear justifications as to which summary statistic should be looked at. This study also gives an alternative interpretation and a simple test statistic for that framework regarding the same null hypothesis.

2. Theory

(a) Preliminaries

Let us first 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

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}, \tag{2.1}$$

where $\bar{x}$ and $\bar{y}$ are the sample means: $\bar{x} = \sum_{i=1}^{N} x_i/N$, $\bar{y} = \sum_{i=1}^{N} y_i/N$. By using the matrix notation

$\mathbf{x} = (x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_N - \bar{x})^\mathrm{T}$ and $\mathbf{y} = (y_1 - \bar{y}, y_2 - \bar{y}, \ldots, y_N - \bar{y})^\mathrm{T}$, where the superscript T denotes transpose, we can rewrite (2.1) as

$$r = \frac{\mathbf{x}^\mathrm{T}\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}, \tag{2.2}$$

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

$$\mathbf{x}^\mathrm{T}\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\| \cos\theta, \tag{2.3}$$

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

$$r = \cos\theta. \tag{2.4}$$

That is, the correlation coefficient and the angle between 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, $r = 0$, the two vectors are perpendicular to each other, $\theta = \pi/2$.

We could standardize the variables by their standard deviations beforehand: $\mathbf{u} = \|\mathbf{x}\|^{-1}\mathbf{x}$ and $\mathbf{v} = \|\mathbf{y}\|^{-1}\mathbf{y}$, so that

$$r = \cos\theta = \mathbf{u}^\mathrm{T}\mathbf{v}. \tag{2.5}$$

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. 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. [49,50]). For what follows, it is convenient to consider the latter with the population mean 0.

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

(b) Distribution of random angles

Let us consider a pair of random vectors $\mathbf{x} = (x_1, x_2, \ldots, x_k)^\mathrm{T}$ and $\mathbf{y} = (y_1, y_2, \ldots, y_k)^\mathrm{T}$ and the angle θ between them. The elements are assumed to be independently and identically distributed. Let $b = (\mathbf{y}^\mathrm{T}\mathbf{x})/(\mathbf{x}^\mathrm{T}\mathbf{x})$, 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 (2.3), 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 (figure 1e). 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

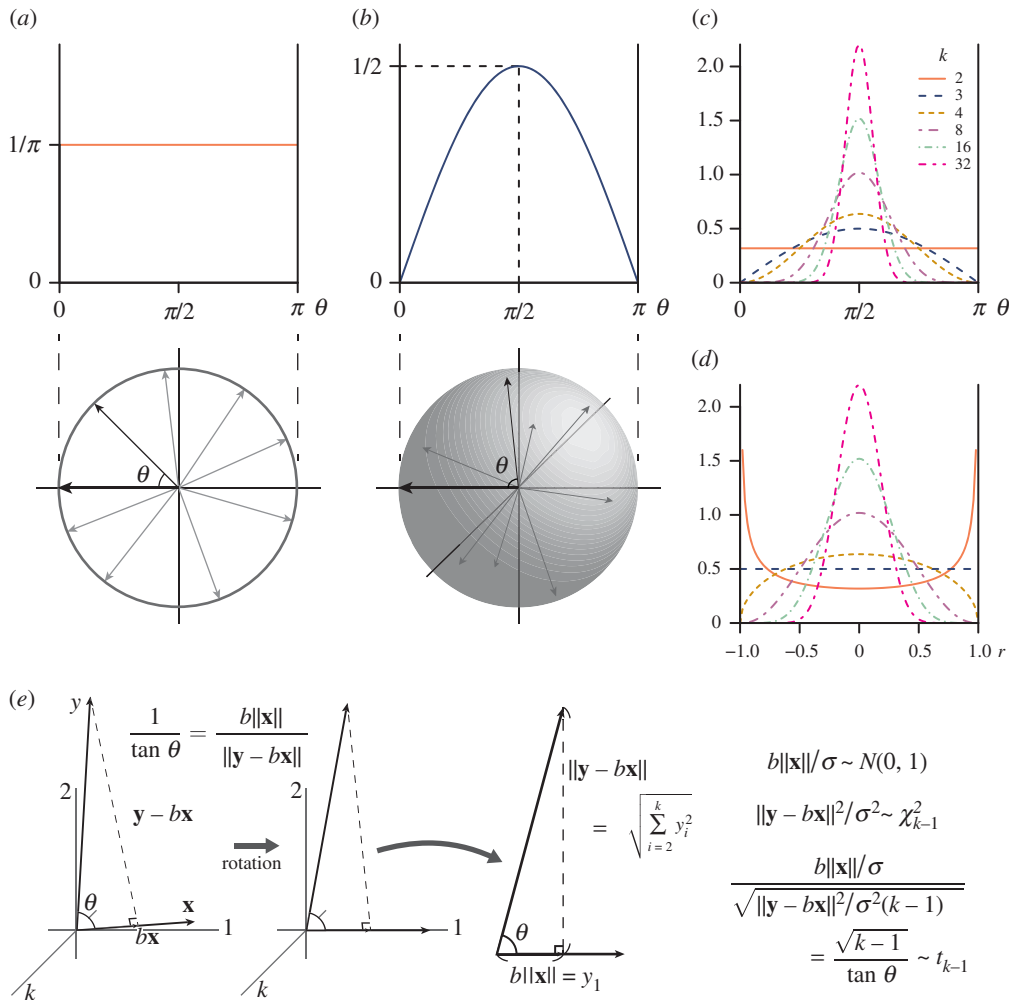$$\frac{1}{\tan\theta} = \frac{b\|\mathbf{x}\|}{\|\mathbf{y} - b\mathbf{x}\|}. \tag{2.6}$$

Figure 1. Distribution of angle in multidimensional spaces. (a,b) Probability density of angle θ between two vectors uniformly distributed on two-dimensional circle and three-dimensional sphere, respectively. Lower panels show schematic illustrations of the angle between a 'pole' (the thick arrow pointing the left-hand side) and another vector uniformly distributed on the unit circle/sphere. Upper panels show the corresponding densities. (c) Density of θ in general k-dimensional cases (2.9). (d) Density of $r = \cos\theta$ (2.8). (e) 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.

The distribution of this quantity is heuristically derived here; see, e.g. [49–52] for formal proofs. Assume 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 standardized 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 the distribution of θ is concerned; 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 χ^2 with $k-1$ degrees of freedom, and independent of $b\|\mathbf{x}\|$ (figure 1). 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 χ^2 variate divided by its degrees of freedom, with the two variates independent from each other—the quantity

$$\frac{b\|\mathbf{x}\|/\sigma}{\sqrt{\|\mathbf{y}-b\mathbf{x}\|^2/\sigma^2(k-1)}} = \frac{\sqrt{k-1}}{\tan\theta} = \sqrt{k-1}\frac{r}{\sqrt{1-r^2}} \qquad (2.7)$$

has a t-distribution with $k-1$ degrees of freedom. The probability density (or element) of r in this case can be derived by

transforming that of the t-distribution:

$$\frac{1}{B(1/2, (k-1)/2)}(1-r^2)^{(k-3)/2}\,dr, \quad -1 \le r \le 1, \qquad (2.8)$$

where $B(a, b)$ is the beta function with the two parameters a and b (this is just a normalizing constant, whose value need not concern most readers) (figure 1d). Then the density for $\theta = \arccos r$ is, by noting $|dr| = |-\sin\theta\,d\theta|$,

$$\frac{1}{B(1/2, (k-1)/2)}(1-\cos^2\theta)^{(k-3)/2}|-\sin\theta\,d\theta|$$
$$= \frac{1}{B(1/2, (k-1)/2)}\sin^{k-2}\theta\,d\theta, \quad 0 \le \theta \le \pi. \qquad (2.9)$$

This density has a peak at $\theta = \pi/2$, which is increasingly pronounced as k increases (figure 1c). Another useful expression can be derived for $s = r^2$, by noting $|dr| = |ds/2\sqrt{s}|$ and duplication of the positive and negative branches for r in (2.8):

$$\frac{1}{B(1/2, (k-1)/2)}s^{-1/2}(1-s)^{(k-3)/2}\,ds, \quad 0 \le s \le 1, \qquad (2.10)$$

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 as long as it is independent of $\mathbf{y}$ that in turn has a spherically contoured distribution [50,51]. Indeed, expressions equivalent to (2.9) and (2.10) can be obtained from purely geometric evaluation of the surface area of a hyperspherical cap [44,53], which is equivalent to the probability for a random vector uniformly distributed on the hypersphere to fall within the region (see also [46]). A similar geometric reasoning was in fact involved in Fisher's [54,55] formal derivation of the t-distribution (see also [56, ch. 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 (2.7); example functions for the R environment [57] are provided in the electronic supplementary 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 (2.10) can be used instead. An equivalent test is commonly used for testing differences between allometric axes (e.g. [45]).

(c) 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. When interest is in analysing a set of many lineages simultaneously (e.g. [37,40,41]), a convenient procedure 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 standardized so that each row has the length of unity:

$$\mathbf{Z} = \text{diag}(\|\mathbf{x}_i\|^{-1})\mathbf{X}, \tag{2.11}$$

where $\text{diag}(\cdot)$ 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

$$\mathbf{C} = \mathbf{Z}\mathbf{Z}^T. \tag{2.12}$$

By construction, $\mathbf{C}$ is symmetric and its (i, j)th elements are the vector correlations between the ith and jth vectors (2.5), with the diagonal elements being 1. The rows need not be centred, and 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 the 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 pairwise angles and conducting a test of locations, e.g. t-test, Wilcoxon rank-sum test. However, such inferences should be, if at all, made with caution, because pairwise angles and correlations are generally not independent from one another. Ordinary statistical tests 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 non-zero covariances unless $\mathbf{\Gamma} = \mathbf{I}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix [58,59]. Similar should be the case for pairwise angles. Therefore, it is inadvisable to conduct tests for pairwise angles in this way, unless, perhaps, the covariances are appropriately taken into account (methods for which are available for correlations; [60,61]). Although a sensible Monte Carlo design could be constructed to accommodate the covariances, it is rather questionable whether tests on mean pairwise angles are of much practical use beyond testing the null hypothesis $\mathbf{\Gamma} = \mathbf{I}_n$. There are more straightforward ways to test this null hypothesis (below), and other cases hardly translate into particular values of mean pairwise angles.

(d) Eigenanalysis and one-step test for multiple vectors

De Lisle & Bolnick [48] 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 $\mathbf{C}$:

$$\mathbf{C} = \mathbf{U}\mathbf{L}\mathbf{U}^T, \tag{2.13}$$

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 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:

$$\mathbf{A} = \mathbf{Z}^T\mathbf{Z} = \mathbf{V}\mathbf{K}\mathbf{V}^T, \tag{2.14}$$

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 (electronic supplementary material, 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 (electronic suupplementary material, appendix A). The rest of this section addresses quantification and test of the magnitude of parallelism—the first objective of the eigenanalysis as originally proposed [48]. Brief comments on the second objective—determination of dimensionality of parallel trajectories—are given in appendix A.

One complexity in dealing with eigenvalues of these matrices is the presence of sampling error and bias, which render sample eigenvalues inaccurate estimators of the corresponding population eigenvalues (e.g. [62,63]). Regarding the null hypothesis test of no parallelism, it has been suggested to compare eigenvalues of $\mathbf{C}$ with Monte Carlo distributions of eigenvalues of matrices drawn from a Wishart distribution [48]. To be clear, that distribution pertains to unscaled cross-product matrices, so the generated random matrices should be scaled as correlation matrices (this scaling was not clearly mentioned in DeLisle & Bolnick's descriptions, although was involved in their computer codes). Although this procedure is potentially valid, it has not been clearly

indicated which test statistic should be looked at in testing this null hypothesis.

Here, it is proposed that dispersion of eigenvalues in these matrices, or equivalently sum of squared correlation coefficients from $\mathbf{C}$, can be an appropriate test statistic. Eigenvalue dispersion has been used to quantify covariation between traits [64–67], and its sampling properties are relatively well known [68]. Intuitively, if phenotypic change vectors are uniformly distributed in the trait space, eigenvalues of $\mathbf{A}$ (or $\mathbf{C}$) exhibit low dispersion. If the vectors are concentrated in a single or a few directions, then the eigenvalues are highly dispersed.

In particular, it is possible to show the following equality regarding dispersions of eigenvalues of $\mathbf{C}$ and $\mathbf{A}$ (denoted l_i and k_i, respectively) and sum of squared correlations (see appendix A):

$$\sum_{i=1}^{n}(l_i - \bar{l})^2 = \sum_{i=1}^{p}(k_i - \bar{k})^2 + \frac{n^2}{p} - n = 2\sum_{i<j}^{n} r_{ij}^2, \qquad (2.15)$$

where $\bar{l}$ and $\bar{k}$ are the averages of eigenvalues, and r_{ij} are the (i, j)th elements of $\mathbf{C}$. Under the null hypothesis that all vectors are independently directed from one another without preferred directions, the population (true) correlation coefficients are zero, or $\mathbf{\Gamma} = \mathbf{I}_n$. For n lineages, we take it 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$[1/2, (p-1)/2]$ (2.10) 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 [68,69]. Therefore, the expectation and variance of the sum of squared correlations are:

$$\mathrm{E}\left(\sum_{i<j}^{n} r_{ij}^2\right) = \frac{n(n-1)}{2p} \quad \text{and} \quad \mathrm{Var}\left(\sum_{i<j}^{n} r_{ij}^2\right) = \frac{n(n-1)(p-1)}{p^2(p+2)}. \qquad (2.16)$$

From these moments, Schott [69] proposed the following high-dimensional asymptotic test for the hypothesis $\mathbf{\Gamma} = \mathbf{I}_n$. 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 $\lim(\mathrm{Var}(\sum_{i<j}^{n} r_{ij}^2)) = \gamma^2$. (Note that this condition is just a modest generalization 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 (2.16), and a large deviation can be seen as evidence against the null hypothesis, suggesting concentration of vectors. Schott [69] 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. less than 16, in which case Monte Carlo simulations can be used) and a power usually superior to that of the conventional likelihood-ratio test.

A caveat on this procedure is that the test statistic does not convey information on the signs of correlation coefficients. Therefore, this test does not distinguish unimodal and antipodal concentration patterns (neither do tests entirely based on eigenvalues). It is strongly recommended to inspect $\mathbf{C}$ or PC scores to surmise what type of deviation from the null is present (see below). If the detection of parallel signal is of specific interest, it is probably more adequate to use the Rayleigh test from the directional statistics (electronic supplementary material, appendix B).

3. Recommendations

Although the statistical toolkits described above enable tests of particular null hypotheses, it is strongly recommended to conduct exploratory analyses before those tests are applied, in order to surmise overall patterns in the data. A common option is to visualize metric relationships between ancestral and descendant states via an ordination method like principal component analysis (PCA). Complementary to this approach is to make ordination of phenotypic change vectors, as is done for allometric axes [12,13]. The latter can be obtained from the eigenanalysis of $\mathbf{A}$ (2.14). It would also be useful to visualize relationships between ordination axes and traits via biplot, or to explore potential structures with clustering approaches (see electronic supplementary material, appendix C).

The Schott and Rayleigh tests share the same null hypothesis that the directional vectors are uniformly distributed, but have different alternative hypotheses. The Rayleigh test is powerful in detecting unimodal concentration of the vectors, but will be senseless if the vectors show antipodal or girdle-like distributions. The Schott test can detect these forms of deviation from uniformity, but does not distinguish antipodal and unimodal patterns by itself. Choice between these different tests should be made according to their properties and biological/statistical hypotheses of interest.

4. Example analysis

Stuart et al.'s [37] 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 electronic supplementary material, appendix D. The resultant dataset consists of 13 phenotypic change vectors in 80 nominal morphological traits: 41 linear measurements, 38 Procrustes-aligned shape coordinates and one centroid size (from two-dimensional geometric morphometric analysis of 19 full landmarks). The effective dimensionality of the vectors is $80 - 4 = 76$, as 4 degrees of freedom are lost by Procrustes alignment (assuming that the configurations were projected onto the tangent space).

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 (2.9), 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 error rate control is deemed necessary for this demonstrative analysis; figure 2a). 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; figure 2b). This interpretation is in stark contrast with that of Stuart et al. [37], who regarded their mean of $81.1°$ with 84 traits as 'nearly orthogonal'. Note, however, that this test is for illustrative purposes only, as the mean pairwise angle lacks a clear interpretability (see above).

Ordination from PCA of the standardized phenotypic change vectors $\mathbf{Z}$ is shown in figure 2c–f (see electronic supplementary material, appendix C for details). The vectors of
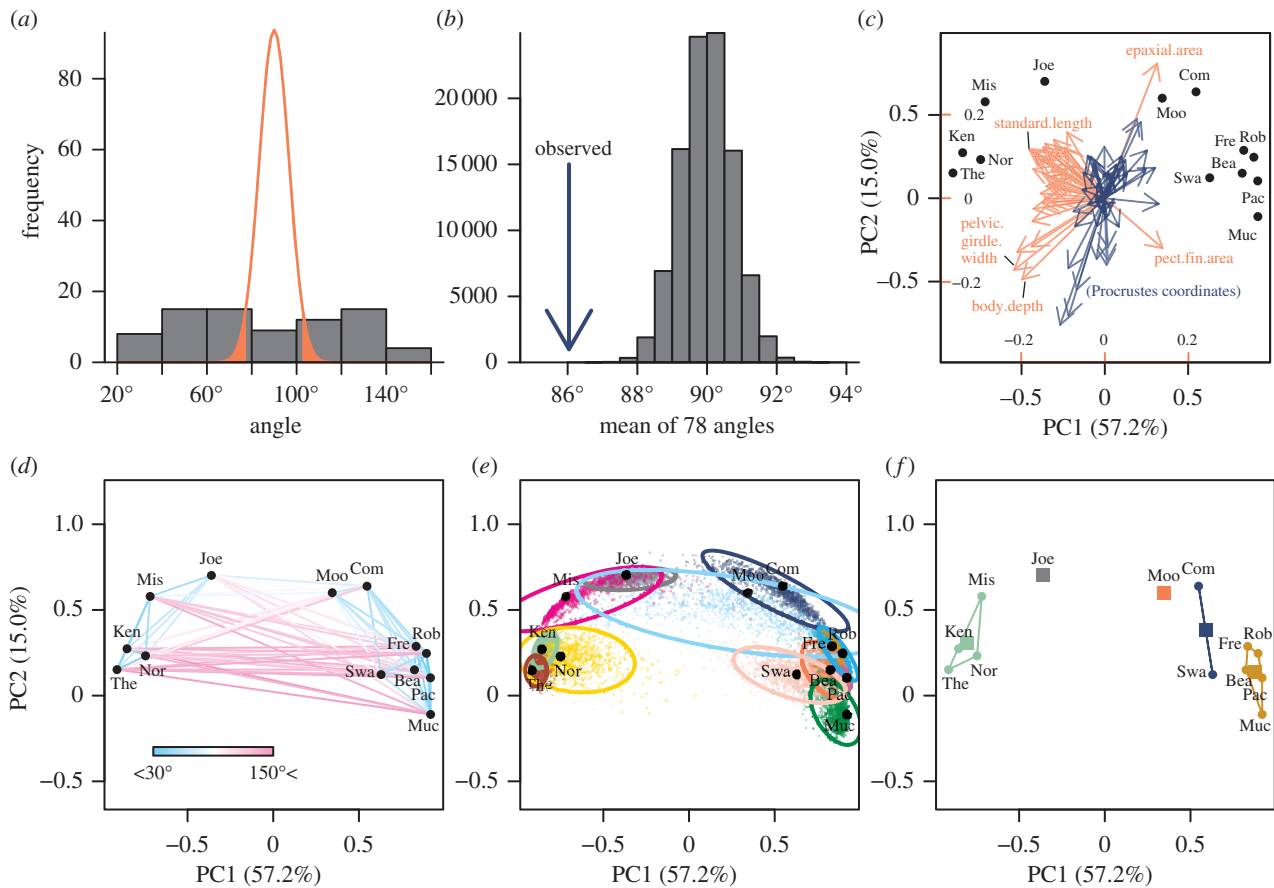
Figure 2. Re-analysis of Stuart *et al.*'s [37] dataset. (*a*) Histogram of 78 pairwise angles between the phenotypic change vectors in 13 lineages of *Gasterosteus aculeatus*, compared with a scaled density of random angles for $p = 76$ (2.9). Regions outside the 2.5 and 97.5 percentiles of the density 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*) PC plots of phenotypic change vectors with different visualizations. (*c*) PCA biplot showing scores (points) and coefficients (arrows) of PC1 and PC2. The scaling parameter α was set to 1 (see electronic supplementary material, appendix C). Blue arrows denote 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. Acronyms for watersheds are as in [37].

some lineages appear closely clustered with one another, but distribution of PC scores across the origin indicates that not all lineages had similar divergence (figure 2*d*). Non-parametric bootstrapping suggests that differences between vectors are mostly larger than what would be expected from sampling error alone except in most similar pairs (figure 2*e*). Nevertheless, the magnitude of sampling error appears heterogeneous among lineages, cautioning against face-value interpretation of differences; for example, sampling error in direction is evidently large for the Moore watershed, and this seems largely owing to small trajectory length. Potential clustering was explored with *k*-means clustering with varying numbers of clusters, and the result for $k = 5$ is shown as an example (figure 2*f*). PCA biplot shows that major components of variation among trajectories are to some extent characterized by variation in standard length and other traits highly correlated with it, along with several others (figure 2*c*). Overall, these exploratory analyses suggest the presence of multiple preferred directions of phenotypic change vectors.

In order to show deviation from the uniformity, the Schott test was applied to this dataset. The sum of the 78 squared correlations was 24.25, whereas the null expectation and standard deviation (from (2.16) with $n = 13$ and $p = 76$) were 1.03 and 0.16, respectively, indicating a statistically significant deviation

from the null hypothesis of uniformity ($Z = 144.08$; $p < 1 \times 10^{-10}$). This test and examinations of pairwise angles and PC scores altogether indicate that the phenotypic change vectors most likely have preferred directions in the trait space, and that some of the vectors are significantly more (dis)similar to one another than expected from uniform distribution by chance. This insight is in contrast to the original account [37], and is partly in line with the reanalysis in [48], reinforced with more rigorous statistics. These results provide objective justifications to explore potential biological causes of the perceived patterns. In particular, the presence of multiple clusters may potentially reflect differing evolvability or selection regimes among recognized clusters. Such possibilities would deserve a more inclusive approach as was originally undertaken by Stuart *et al.* [37], with the aid of the present methodology.

5. Discussion

Angles have been commonly used in quantitative analyses of parallel evolution, but their properties in multidimensional spaces have not attained due attention. As clarified by the above analysis, angles between randomly directed vectors

are peaked around the right angle in multidimensional spaces ((2.9); figure 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 [33,37]. In addition, the dependency of the peakedness on dimensionality ((2.9); figure 1c) renders angles incomparable across different dimensions. Thus, direct comparison of angles or pooled meta-analysis across varying dimensionalities [33,38,70] will be tenuous, unless dimensionality is sensibly taken into account. A potentially useful standardization in this respect is $\sqrt{k-2}(\pi/2-\theta)$, whose distribution under the null condition (2.9) converges to the standard normal distribution as $k \to \infty$ [52]; when k is sufficiently large, this quantity could be used as an effect size against the null distribution.

This review has 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 [7,48]. This is not to claim superior biological importance of this hypothesis over another, but rather to present it as a baseline for analysing multidimensional vectors. At the other extreme, the hypothesis of completely parallel vectors could be tested, if interest is in detecting deviation from parallelism [7]. 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 correlation in the population eliminates any room for sampling variation, thereby trivially yielding sample correlation coefficient exactly 1 or −1 with probability 1. (On the other hand, tests against no preferred directions described above are not seriously affected by sampling error, with which the uniform distribution on the hypersphere typically remains unaffected under the null hypothesis [71].) A more realistic option will be to adopt one of the resampling-based approaches [3,20,32,72], as is done in the phenotypic trajectory analysis. However, it should be remembered that a resampling-based test, although being nominally non-parametric, is usually not free from the assumption that the populations share the same form of distribution, potentially differing only in the quantity of interest [73,74]. Between-group heteroscedasticity, whose presence was also suggested in the present re-analysis (figure 2e), can possibly undermine adequacy of tests of this type. Robustness of resampling-based tests against such cases needs to be critically assessed.

Apart from hypothesis testing, exploratory methods could be more commonly used in the analysis of parallelism. 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 [14,24], test for shared trajectories in subspaces [75–77], and simultaneous visualization with phylogeny [15]. Although not fully discussed here, clustering approaches may also be useful in detecting patterns in multiple phenotypic change vectors (figure 2f; electronic supplementary material, appendix C).

It may be worth emphasizing that all statistical techniques described here fundamentally pertain to patterns of observed data rather than biological causalities. It is widely recognized that parallel and convergent evolutionary patterns (or lack thereof) can arise from a number of disparate causes [7,9,78]. Associating statistical techniques to any of such possibilities is restrained here, as it can lead to conflation between patterns and causalities and/or biased interpretation of statistical results [2]. It should be remembered that inferences for biological causality typically require more inclusive approaches than analysis of evolutionary patterns alone [79,80].

A paramount assumption in almost any geometric analysis of trajectories in a trait space [2,3,7] is that vectors can be meaningfully compared across different regions of the trait space. For angles to be meaningful, the space needs to be Euclidean [81,82]. This is, for example, when all traits are measured in the same unit. If traits are in different units (e.g. linear measurements and mass), the magnitude of measured angles can vary depending on the arbitrary choice of units. Log-transformation or standardization by mean or standard deviation would make traits nominally dimensionless, but it is generally an open question whether they ensure interpretability of vectors and angles.

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 [83,84]. Biological interpretations of evolutionary trajectories in high-dimensional trait spaces should be underpinned by proper understanding of the relevant geometry and statistics. In this regard, the literature of directional statistics (e.g. [71,85,86]) may potentially provide useful directions for future methodological developments.

References

1. Stayton CT. 2008 Is convergence surprising? An examination of the frequency of convergence in simulated datasets. *J. Theor. Biol.* **252**, 1–14. (doi:10.1016/j.jtbi.2008.01.008)

2. Stayton CT. 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)

3. Adams DC, Collyer ML. 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)

4. Adams DC, Collyer ML. 2019 Phylogenetic comparative methods and the evolution of multivariate phenotypes. *Annu. Rev. Ecol. Evol. Syst.* **50**, 405–425. (doi:10.1146/annurev-ecolsys-110218-024555)

5. Arbuckle K, Bennett CM, Speed MP. 2014 A simple measure of the strength of convergent evolution. *Methods Ecol. Evol.* **5**, 685–693. (doi:10.1111/2041-210X.12195)

6. Speed MP, Arbuckle K. 2017 Quantification provides a conceptual basis for convergent evolution. *Biol. Rev.* **92**, 815–829. (doi:10.1111/brv.12257)

7. Bolnick DI, Barrett RDH, Oke KB, Rennison DJ, Stuart YE. 2018 (Non)parallel evolution. *Annu. Rev. Ecol. Evol. Syst.* **49**, 303–330. (doi:10.1146/annurev-ecolsys-110617-062240)

8. Stayton CT. 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)

9. Agrawal AA. 2017 Toward a predictive framework for convergent evolution: integrating natural history, genetic mechanisms, and consequences for the diversity of life. *Am. Nat.* **190**, S1–S12. (doi:10.1086/692111)

10. Blount ZD, Lenski RE, Losos JB. 2018 Contingency and determinism in evolution: replaying life's tape. *Science* **362**, eaam5979. (doi:10.1126/science.aam5979)

11. Rincon-Sandoval M *et al.* 2020 Evolutionary determinism and convergence associated with water-column transitions in marine fishes. *Proc. Natl Acad. Sci. USA* **117**, 33 396–33 403. (doi:(10.1073/pnas.2006511117))

12. Solignac M, Cariou ML, Wimitzky M. 1990 Variability, specificity and evolution of growth gradients in the species complex *Jaera albifrons* (Isopoda, Asellota). *Crustaceana* **59**, 121–145. (doi:10.1163/156854090X00615)

13. Klingenberg CP, Froese R. 1991 A multivariate comparison of allometric growth patterns. *Syst. Zool.* **40**, 410–419. (doi:10.1093/sysbio/40.4.410)

14. Gerber S, Eble GJ, 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)

15. Baliga VB, Mehta RS. 2018 Phylo-allometric analyses showcase the interplay between life-history patterns and phenotypic convergence in cleaner wrasses. *Am. Nat.* **191**, E129–E143. (doi:10.1086/697047)

16. Strelin MM, Benitez-Vieyra S, Fornoni J, Klingenberg CP, Cocucci A. 2018 The evolution of floral ontogenetic allometry in the Andean genus *Caiophora* (Loasaceae, subfam. Loasoideae). *Evol. Dev.* **20**, 29–39. (doi:10.1111/ede.12246)

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

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

19. Shea BT. 1985 Bivariate and multivariate growth allometry: statistical and biological considerations. *J. Zool. A* **206**, 367–390. (doi:10.1111/j.1469-7998.1985.tb05665.x)

20. Klingenberg CP. 1996 Multivariate allometry. In *Advances in morphometrics* (ed. LF Marcus), pp. 23–49. New York, NY: Plenum Press.

21. Klingenberg CP. 2016 Size, shape, and form: concepts of allometry in geometric morphometrics. *Dev. Genes Evol.* **226**, 113–137. (doi:10.1007/s00427-016-0539-2)

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

23. Zelditch ML, Calamari ZT, Swiderski DL. 2016 Disparate postnatal ontogenies do not add to the shape disparity of infants. *Evol. Biol.* **43**, 188–207. (doi:10.1007/s11692-016-9370-y)

24. Urošević A, Ljubisavljević K, Ivanović A. 2013 Patterns of cranial ontogeny in lacertid lizards: morphological and allometric disparity. *J. Evol. Biol.* **26**, 399–415. (doi:10.1111/jeb.12059)

25. Wilson LAB. 2013 Allometric disparity in rodent evolution. *Ecol. Evol.* **3**, 971–984. (doi:10.1002/ece3.521)

26. Wilson LAB. 2018 The evolution of ontogenetic allometric trajectories in mammalian domestication. *Evolution* **72**, 867–877. (doi:10.1111/evo.13464)

27. Rohner PT. 2020 Evolution of multivariate wing allometry in schizophoran flies (Diptera: Schizophora). *J. Evol. Biol.* **33**, 831–841. (doi:10.1111/jeb.13613)

28. Pérez-Ben CM, Báez AM, Schoch RR. 2020 Morphological evolution of the skull roof in temnospondyl amphibians mirrors conservative ontogenetic patterns. *Zool. J. Linn. Soc.* **188**, 163–179. (doi:10.1093/zoolinnean/zlz068)

29. Feiner N, Jackson ISC, Van der Cruyssen E, Uller T. 2021 A highly conserved ontogenetic limb allometry and its evolutionary significance in the adaptive radiation of *Anolis* lizards. *Proc. R. Soc. B* **288**, 20210226. (doi:10.1098/rspb.2021.0226)

30. Adams DC, Collyer ML. 2007 Analysis of character divergence along environmental gradients and other covariates. *Evolution* **61**, 510–515. (doi:10.1111/j.1558-5646.2007.00063.x)

31. Collyer ML, Adams DC. 2007 Analysis of two-state multivariate phenotypic change in ecological studies. *Ecology* **88**, 683–692. (doi:10.1890/06-0727)

32. Collyer ML, Adams DC. 2013 Phenotypic trajectory analysis: comparison of shape change patterns in evolution and ecology. *Hystrix* **24**, 75–83. (doi:10.4404/hystrix-24.1-6298)

33. Oke KB, Rolshausen G, LeBlond C, Hendry AP. 2017 How parallel is parallel evolution? A comparative analysis in fishes. *Am. Nat.* **190**, 1–16. (doi:10.1086/691989)

34. Kaeuffer R, Peichel CL, Bolnick DI, Hendry AP. 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)

35. Raeymaekers JAM, Chaturvedi A, Hablützel PI, Verdonick I, Hellermans B, Maes GE, De Meester L, Volckaert FAM. 2017 Adaptive and non-adaptive divergence in a common landscape. *Nat. Commun.* **8**, 267. (doi:10.1038/s41467-017-00256-6)

36. Langerhans RB. 2018 Predictability and parallelism of multitrait adaptation. *J. Hered.* **109**, 59–70. (doi:10.1093/jhered/esx043)

37. Stuart YE *et al.* 2017 Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nat. Ecol. Evol.* **1**, 0158. (doi:10.1038/s41559-017-0158)

38. Haines GE, Stuart YE, Hanson D, Tasneem T, Bolnick DI, Larsson HCE, Hendry AP. 2020 Adding the third dimension to studies of parallel evolution of morphology and function: an exploration based on parapatric lake-stream stickleback. *Ecol. Evol.* **10**, 13 297–13 311. (doi:10.1002/ece3.6929)

39. Jacobs A *et al.* 2020 Parallelism in eco-morphology and gene expression despite variable evolutionary and genomic backgrounds in a Holarctic fish. *PLoS Genet.* **16**, e1008658. (doi:10.1371/journal.pgen.1008658)

40. James ME, Wilkinson MJ, Bernal DM, Liu H, North HL, Engelstädter J, Ortiz-Barrientos D. 2021 Phenotypic and genotypic parallel evolution in parapatric ecotypes of *Senecio*. *Evolution* **75**, 3115–3131. (doi:10.1111/evo.14387)

41. Owens GL, Veen T, Moxley DR, Arias-Rodriguez L, Tobler M, Renninson DJ. 2022 Parallel shifts of visual sensitivity and body coloration in replicate populations of extremophile fish. *Mol. Ecol.* **31**, 946–958. (doi:10.1111/mec.16279)

42. Weber AAT, Rajkov J, Smailus K, Egger B, Salzburger W. 2021 Speciation dynamics and extent of parallel evolution along a lake–stream environmental contrast in African cichlid fishes. *Sci. Adv.* **7**, eabg5391. (doi:10.1126/sciadv.abg5391)

43. de Aranzamendi MC, Martínez JJ, Held C, Sahade R. 2022 Parallel shape divergence between ecotypes of the limpet *Nacella concinna* along the Antarctic Peninsula: a new model species for parallel evolution? *Zoology* **150**, 125983. (doi:10.1016/j.zool.2021.125983)

44. Rice SH. 1990 A geometric model for the evolution of development. *J. Theor. Biol.* **143**, 319–342. (doi:10.1016/S0022-5193(05)80033-5)

45. Klingenberg CP, Marugán-Lobón J. 2013 Evolutionary covariation in geometric morphometric data: analysing integration, modularity, and allometry in a phylogenetic context. *Syst. Biol.* **62**, 591–610. (doi:10.1093/sysbio/syt025)

46. Ram Y, Hadany L. 2015 The probability of improvement in Fisher's geometric model: a probabilistic approach. *Theor. Popul. Biol.* **99**, 1–6. (doi:10.1016/j.tpb.2014.10.004)

47. Thompson KA, Osmond MM, Schluter D. 2019 Parallel genetic evolution and speciation from standing variation. *Evol. Lett.* **3**, 129–141. (doi:10.1002/evl3.106)

48. De Lisle SP, Bolnick DI. 2020 A multivariate view of parallel evolution. *Evolution* **74**, 1466–1481. (doi:10.1111/evo.14035)

49. Hotelling H. 1953 New light on the correlation coefficient and its transforms. *J. R. Stat. Soc. B* **15**, 193–232. (doi:10.1111/j.2517-6161.1953.tb00135.x)

50. Anderson TW. 2003 *An introduction to multivariate statistical analysis*, 3rd edn. Hoboken, NJ: John Wiley & Sons.

51. Muirhead RJ. 1982 *Aspects of multivariate statistical theory*. Hoboken, NJ: John Wiley & Sons.

52. Cai T, Fan J, Jiang T. 2013 Distributions of angles in random packing on spheres. *J. Mach. Learn. Res.* **14**, 1837–1864.

53. Li S. 2011 Concise formulas for the area and volume of a hyperspherical cap. *Asian J. Math. Stat.* **4**, 66–70. (doi:10.3923/ajms.2011.66.70)

54. Fisher RA. 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)

55. Fisher RA. 1925 Applications of "Student's" distribution. *Metron* **5**, 90–104.

56. Stuart A, Ord JK. 1994 *Kendall's advanced theory of statistics*, vol. 1, 6th edn. London, UK: Hodder Education. [Reprinted in 2004 by John Wiley & Sons, Chichester, UK.]

57. R Core Team. 2019 *R: a language and environment for statistical computing, version 3.5.3*. Vienna, Austria: R Foundation for Statistical Computing. See https://www.R-project.org/.

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

59. Olkin I, Finn J. 1990 Testing correlated correlations. *Psych. Bull.* **108**, 330–333. (doi:10.1037/0033-2909.108.2.330)

60. Olkin I, Finn JD. 1995 Correlations redux. *Psych. Bull.* **118**, 155–164. (doi:10.1037/0033-2909.118.1.155)

61. Zou GY. 2007 Toward using confidence intervals to compare correlations. *Psych. Bull.* **12**, 399–413. (doi:10.1037/1082-989X.12.4.399)

62. Anderson TW. 1963 Asymptotic theory for principal component analysis. *Ann. Math. Stat.* **34**, 122–148. (doi:10.1214/aoms/1177704248)

63. Jolliffe IT. 2002 *Principal component analysis*, 2nd edn. New York, NY: Springer.

64. Cheverud JM, Rutledge JJ, Atchley WR. 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)

65. Wagner GP. 1984 On the eigenvalue distribution of genetic and phenotypic dispersion matrices: evidence for a nonrandom organization of quantitative character variation. *J. Math. Biol.* **21**, 77–95. (doi:10.1007/BF00275224)

66. Pavlicev M, Cheverud JM, Wagner GP. 2009 Measuring morphological integration using eigenvalue variance. *Evol. Biol.* **36**, 157–170. (doi:10.1007/s11692-008-9042-7)

67. Haber A. 2011 A comparative analysis of integration indices. *Evol. Biol.* **38**, 476–488. (doi:10.1007/s11692-011-9137-4)

68. Watanabe J. 2022 Statistics of eigenvalue dispersion indices: quantifying the magnitude of phenotypic integration. *Evolution* **76**, 4–28. (doi:10.1111/evo.14382)

69. Schott JR. 2005 Testing for complete independence in high dimensions. *Biometrika* **92**, 951–956. (doi:10.1093/biomet/92.4.951)

70. Radersma R, Noble DWA, Uller T. 2020 Plasticity leaves a phenotypic signature during local adaptation. *Evol. Lett.* **4**, 360–370. (doi:10.1002/evl3.185)

71. Mardia KV, Jupp PE 1999 *Directional statistics*. Chichester, UK: John Wiley & Sons.

72. Sheets HD, Zelditch ML. 2013 Studying ontogenetic trajectories using resampling methods and landmark data. *Hystrix* **24**, 67–73. (doi:10.4404/hystrix-24.1-6332)

73. Anderson MJ. 2001 A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**, 32–46. (doi:10.1111/j.1442-9993.2001.01070.pp.x)

74. Manly BFJ, Navarro Alberto JAJ. 2020 *Randomization, bootstrap and Monte Carlo methods in biology*, 4th edn. Boca Raton, FL: Chapman and Hall.

75. Klingenberg CP. 1998 Heterochrony and allometry: the analysis of evolutionary change in ontogeny. *Biol. Rev.* **73**, 79–123. (doi:10.1017/S000632319800512X)

76. Mitteroecker P, Gunz P, Bookstein FL. 2005 Heterochrony and geometric morphometrics: a comparison of cranial growth in *Pan paniscus* versus *Pan troglodytes*. *Evol. Dev.* **7**, 244–258. (doi:10.1111/j.1525-142X.2005.05027.x)

77. Gerber S, Neige P, Eble GJ. 2007 Combining ontogenetic and evolutionary scales of morphological disparity: a study of early Jurassic ammonites. *Evol. Dev.* **9**, 472–482. (doi:10.1111/j.1525-142X.2007.00185.x)

78. Losos JB. 2011 Convergence, adaptation, and constraint. *Evolution* **65**, 1827–1840. (doi:10.1111/j.1558-5646.2011.01289.x)

79. Losos JB. 2011 Seeing the forest for the trees: the limitations of phylogenies in comparative biology. *Am. Nat.* **177**, 709–727. (doi:10.1086/660020)

80. Gardner JD, Organ CL. 2021 Evolutionary sample size and consilience in phylogenetic comparative analysis. *Syst. Biol.* **70**, 1061–1075. (doi:10.1093/sysbio/syab017)

81. Mitteroecker P, Huttegger SM. 2009 The concept of morphospaces in evolutionary and developmental biology: mathematics and metophors. *Biol. Theory* **4**, 54–67. (doi:10.1162/biot.2009.4.1.54)

82. Huttegger SM, Mitteroecker P. 2011 Invariance and meaningfulness in phenotype spaces. *Evol. Biol.* **38**, 335–351. (doi:10.1007/s11692-011-9123-x)

83. Blum A, Hopcroft J, Kannan R. 2020 *Foundations of data science*. Cambridge, UK: Cambridge University Press.

84. Rohlf FJ. 2021 Why clusters and other patterns can seem to be found in analyses of high-dimensional data. *Evol. Biol.* **48**, 1–16. (doi:10.1007/s11692-020-09518-6)

85. Ley C, Verdebout T. 2017 *Modern directional statistics*. Boca Raton, FL: CRC Press.

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