Biological Journal of the Linnean Society, 2004, **83**, 243–260. With 5 figures

Effects of morphometric descriptor changes on statistical classification and morphospaces

NICOLAS NAVARRO[1,*], XAVIER ZATARAIN[2] and SOPHIE MONTUIRE[1,3]

[1]*UMR CNRS 5561 – Biogéosciences, Centre des Sciences de la Terre, Université de Bourgogne, 6 bvd Gabriel, F-21000 Dijon, France*
[2]*Ecole des Mines, 60 bvd St Michel, F-75006 Paris, France*
[3]*EPHE, Université de Bourgogne, 6 bvd Gabriel, F-21000 Dijon, France*

Received 4 April 2003; accepted for publication 26 January 2004

Ten morphometric descriptors (five pairs of form and shape parameters) are used to describe the complex morphology of the first lower molar of two morphologically similar species, *Microtus arvalis* and *M. agrestis*. These descriptors are derived either from linear measurements or from outline analysis. The effects of these different descriptors on classical analysis as used in biology or palaeobiology are explored. First, the reliability of results in statistical classification is assessed. All of the descriptors discriminate well between the two species. The initial morphometric scheme (linear or outline) does not induce marked differences in statistical classification and the major discrepancies are between standardized and non-standardized versions of descriptors, and between amplitude- and coefficient-based or linear-based descriptors. Subsequently, the similarity of morphospaces based on partial least squares analysis and of intraspecific variance (estimated from the morphospace analysis) are observed. This is done within a morphospace-disparity framework and procedures used here for testing are directed at this research area. Similarities between morphospaces are relatively high. In this case, the initial morphometric scheme is a major factor inducing dissimilarity. However, the patterns of intraspecific dispersion inferred from morphospaces are roughly similar. Major differences in results correspond to the two classes of form or shape descriptors. Similarity of intraspecific variance is obtained when standardized descriptors are used (except for amplitude-based descriptors); conversely, dissimilarity is obtained when non-standardized descriptors are used. In many cases, the results of the various analyses are robust despite changes in descriptor. Moreover, the developmental pathway of vole teeth can frequently explain major dissimilarity or even similarity.

ADDITIONAL KEYWORDS: complex series – elliptic series – Fourier – linear measurements – *Microtus* – multivariate analyses – voles.

INTRODUCTION

The aim of a morphometric study is to recover and recombine information in such a way as to obtain the most accurate biological insights. This information may be highly disparate for a number of reasons. The relative amounts of available and unavailable information are dependent on the system of description (e.g. distances, landmarks, outlines), which in turn involves three types of tool (traditional, geometric and outline morphometrics). The structure of a descriptor and the quantity of information contained in it – a combination of parameters obtained from the initial mathematical processing of extracted data – greatly affect the quality and quantity of morphological information.

The efficacy of the tools, or of the processing techniques which those tools comprise, can be compared. However, as the information is not structured in the same way, relationships between descriptor parameters may have varying degrees of complexity. This involves different approaches to the proper constraints of multivariate analyses for the recovery and recombination of the information that is initially available. Ultimately, the information recovered and

*Corresponding author.
E-mail: nicolas.navarro@u-bourgogne.fr

used in subsequent analyses (e.g. evolutionary patterns) may or may not vary between tools and the same tool may produce different results.

Choice of descriptor is a crucial step in any biological or palaeobiological study based on morphology, since all subsequent analyses depend on this and any change can potentially modify the results. A large array of morphometric descriptors (e.g. residuals of Procrustes adjustment, Fourier coefficients, Eigenshape values) are available for processing shape (e.g. Rohlf & Archie, 1984; Rohlf, 1986; Lohmann & Schweitzer, 1990; Marcus *et al*., 1996; Lestrel, 1997a; MacLeod, 1999). The choice of any particular descriptor may be difficult and is often somewhat arbitrary. Traditional descriptors (i.e. linear measurements) are often abandoned in favour of what are considered a priori to be more efficient methods of geometric morphometrics or outline analysis. This is due to the fact that the intrinsic quality of the traditional descriptors varies according to the groups and problems addressed.

Common biological and palaeobiological applications based on morphology are frequently used as a first step in multivariate methods such as discriminant function analysis or principal components analysis (PCA). One common example of discriminant-based techniques is the method whereby a clearly identified extant sample is used in order to make blind determinations of fossil individuals. Another is where the necessary information for correct classification of extant species is unavailable, e.g. particular morphological characters, character associations, genetic information (e.g. Airoldi, Flury & Salvioni, 1995).

One example of PCA-based techniques is the disparity framework used in macroevolutionary studies (e.g. Foote, 1990, 1999; Eble, 2000a, 2002). This methodological framework (morphospace disparity) can easily be extended to microevolutionary studies such as ecophenotypism or species/lineage evolution. The robustness of results *vis-à-vis* descriptor changes (i.e. modification of the accuracy of the taxonomic determination, stability of the observed pattern) will need to be checked, but should allow workers to make an objective choice or at least alert them to the potential for error.

In this study we explore how changes in descriptor affect the results obtained. Our study subjects are the two extant vole species *Microtus arvalis* and *M. agrestis* (Arvicolinae, Rodentia). Arvicolines (voles and lemmings) are a highly diversified Holarctic rodent group comprising 143 extant species (Musser & Carleton, 1993) whose evolutionary radiation began 5.5 Myr ago (Chaline, Brunet-Lecomte & Campy, 1999). Arvicolines, in particular *Microtus*, are relatively well studied in terms of ecology (e.g. Giraudoux *et al*., 1997; Gårding, 2000; Lindström *et al*., 2001), development (e.g. for teeth, Jernvall, Keränen, & Thesleff, 2000; Salazar-Ciudad & Jernvall, 2002), phylogeny (e.g. Conroy & Cook, 2000; Conroy *et al*., 2001), and biogeography (e.g. Conroy, Demboski & Cook, 1999; Jaarola, Tegelström & Fredga, 1999). Moreover, the abundance of arvicoline teeth means there is a relatively well-documented fossil record. This record has proved a rich data source for many palaeontological research programmes, including studies of evolutionary models (Chaline, 1987; Chaline *et al*., 1999) and palaeoclimatology (Chaline *et al*., 1995; Montuire *et al*., 1997).

Applications of statistical species discrimination and morphospace analysis are frequent in this group (Chaline & Laurin, 1986; Brunet-Lecomte, 1988; Chaline *et al*., 1993; Néraudeau *et al*., 1995; Schmittbull *et al*., 1997; Courant *et al*., 1999; Laplana *et al*., 2000; Hurth *et al*., 2004). They are generally based on traditional descriptors corresponding to combinations of linear measurements (Brunet-Lecomte, 1988; Laplana *et al*., 2000). Previous applications to both extant and fossil forms have shown that this descriptor permits discrimination of tiny intraspecific differences (Laplana *et al*., 2000) and that it is a powerful tool for extracting evolutionary information (Chaline & Laurin, 1986; Chaline *et al*., 1993; Néraudeau *et al*., 1995). Geometric morphometrics and outline analysis have more rarely been used on this group (Schmittbull *et al*., 1997; Courant *et al*., 1999; Laplana *et al*., 2000; Hurth *et al*., 2004) but appear to yield similar results.

The present study is an attempt to determine the relative merits of the various morphometric descriptors available. In addition, the method of complex Fourier analysis, which has been seldom used since being introduced to morphometrics in the 1980s, is applied and compared with classical elliptic analysis. Comparisons are made between the descriptors (e.g. statistical classification, morphospace), although no particular case is made for any of them. Our aim is to: (1) quantify the error rate and expected bias for statistical classification; (2) observe similarity of morphospaces and patterns inferred; (3) compare various descriptors with their different intrinsic qualities based on the results obtained.

MATERIAL AND MORPHOMETRIC DESCRIPTORS

SAMPLING

Our analysis was based on the first lower molars of 122 trapped voles from the Frankfurt am Main region (Senckenberg Museum, Germany). The specimens comprise equal numbers of *Microtus arvalis* and *M. agrestis* as identified by criteria other than their

first lower molars (e.g. fur colour, second upper molar morphology, habitats). The first lower molars in these two species display wide morphological variation ranging from highly distinctive extreme forms to intermediate forms that could not readily have been ascribed to either species in the absence of other criteria.

Linear measurement methods

Traditionally, the morphology of first molars in voles is described by a combination of 23 measurements (Fig. 1; Brunet-Lecomte, 1988). This descriptor focuses on the anterior part of the teeth, which has been described as the principal locus of evolutionary modifications in voles in contrast to what is considered a more fixed posterior part (Chaline, 1972). A recent study (Jernvall *et al*., 2000) has reported that tooth development in voles is a process of iterated addition of lateral cusps toward the front and is an example of extreme molarization. Thus, this linear-based morphometric scheme can be viewed as being weighted in favour of morphological features gained in late ontogenetic times.

A second descriptor (log-shape ratio) is a combination of 23 size-standardized and log transformed measurements. Traditionally, tooth length is taken as a proxy for size in arvicoline rodents. However, occlusal surface area seems to provide a better estimate of size or body mass (Creighton, 1980; Gingerich, Smith & Rosenberg, 1982; Legendre, 1989) than tooth length alone. A good estimate of the occlusal surface area with linear measurement appears to be the square root of the sum of all the squared variables (Sundberg, 1996), which measurement gives a closer correlation with occlusal surface area (correlation coefficient: *arvalis* = 0.727; *agrestis* = 0.892; whole sample = 0.923) than tooth length (*arvalis* = 0.685; *agrestis* = 0.867; whole sample = 0.89).

$$LogRatio_i = \log \frac{V_i}{\sqrt{\sum_{i=1}^{23} V_i^2}} \quad (1)$$

Linear measurements are made with a measuroscope and tooth orientation is standardized using a reference axis based on two landmarks located at the base of the first triangle (T1) and the fourth triangle (T4) and positioned on the vertical axis (Fig. 1).

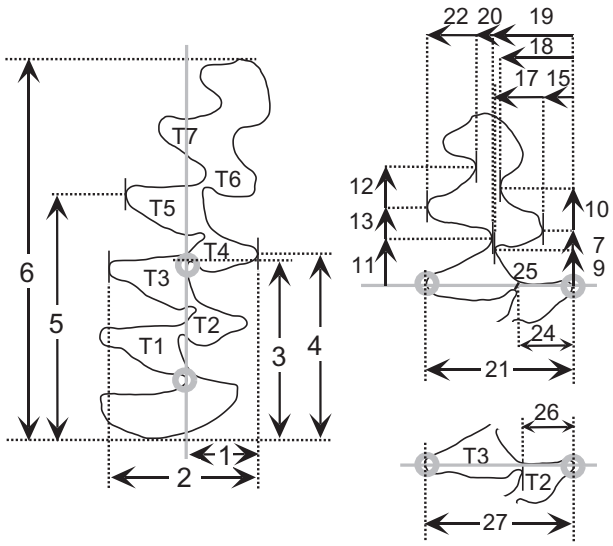

Figure 1. Standardization of orientation and linear measurements used in this study. First lower molars are all positioned so that two landmarks located at the bases of T1 and T4 lie on the vertical axis. This procedure was repeated for outlines after digitization of drawings.

Landmarks

Landmarks-based analysis, i.e. geometric morphometrics, is a powerful tool for characterizing shape. A prerequisite for its use is the precise definition of homologous landmarks. In vole molars, a precise determination of landmarks is relatively difficult and sometimes uncertain. Some landmarks (located in the anterior loops) can disappear, even within the same species. While practicable, geometric morphometrics thus shares the same problems of definition and loss of information that linear morphometrics has. We chose not to include this type of descriptor, preferring traditional methods for characterizing form (linear and outline descriptors) in order to obtain a better characterization.

Outline methods

Vole molars have a very complex outline of loops and triangles. Complex outlines can be accurately described by a number of methods. The most commonly used in morphometrics is elliptic Fourier analysis (EFA; Kuhl & Gardina, 1982; Rohlf & Archie, 1984). A related method, dual-axis Fourier shape analysis (DAFSA; Moellering & Rayner, 1981, 1982), has been used far less since its development (e.g. Kincaid & Schneider, 1983; Bertin *et al*., 2002) despite having the advantage that it can be fitted to a signal in a single pass. Here, we refer to this method as complex discrete Fourier analysis (CDFA).

Common characteristics

Elliptic and complex analyses are similar methods, sharing a number of features in common. Outlines are drawn using a camera lucida. The same procedure is used for standardizing orientation as is employed for

linear measurements (Fig. 1). This was repeated with outlines after digital acquisition of drawings. N equidistant points (here 512) are sampled along a curve. Both types of analysis yield outline decomposition in one or two discrete periodic signals of the *xy* coordinates using a discrete Fourier transform (DFT). Where N is a power of two, DFT uses a Fast Fourier Transform (FFT) algorithm. Otherwise a computationally slower algorithm is used.

$$F(S) = \sum_{\mathrm{k}=0}^{\mathrm{N}-1}\left(\frac{1}{\mathrm{N}}\sum_{j=0}^{N-1} S_j e^{-i\frac{2\pi kj}{N}}\right)e^k \text{ with } \mathrm{e}^{\mathrm{k}} = \mathrm{e}^{+\mathrm{i}\frac{2\pi \mathrm{k}}{\mathrm{N}}} \quad (2)$$

S_j is equal to either a real signal (X_j or Y_j) or a complex signal i.e. $Z_j = X_j + iY_j$ with $j = 0$ to N–1. In each case, N harmonics can be calculated. However, in practice, a peculiar property of conjugate symmetry is that the second half becomes redundant and N/2 harmonics are sufficient. Each harmonic is characterized by its amplitude and phase or by its Fourier coefficients.

$$C_k(S) = \frac{1}{\mathrm{N}}\sum_{j=0}^{N-1} S_j e^{-i\frac{2\pi kj}{N}} \quad (3)$$

$$= A_k(S_j)e^{i\varphi(S_j)} \quad (4)$$

C is the Euler formulation of the Fourier coefficients and its correspondence with cosine-sine formulation is as follows:

$$C_k = a_k + ib_k \quad (5)$$

The amplitude is the modulus of the Fourier coefficient and corresponds to:

$$A_k = \sqrt{a_k^2 + b_k^2} \quad (6)$$

For each descriptor defined by these two outline methods (see below), two versions – size standardized (abbreviated std) and non-standardized (abbreviated nstd) – are considered. Because overall tooth shapes between the two species are so similar, using the area of the occlusal surface as a size proxy is adequate here. Size standardization is obtained from the square root of the area of the occlusal surface.

Elliptic Fourier analysis

This method is widely used in morphometrics (Rohlf & Archie, 1984; Ferson, Rohlf & Koehn, 1985; Lestrel, 1997a; Schmittbull *et al*., 1997; McLellan & Endler, 1998). It employs two real signals, X and Y. Each signal is decomposed into harmonics and N/2 harmonics are obtained (with N = number of sampling points), conjugate symmetry making the other half redundant. Four coefficients are obtained for each harmonic (two for each signal; Eqn 5). This is the first descriptor (EFA C: combination of 4K coefficients, where K = number of harmonics). Two ways of combining amplitude are described in the literature. With the first, the amplitude of each real signal is processed separately and a general formulation is recovered (see Eqn 6), which may be written where there are two signals:

$$\mathrm{A}_k^{\mathrm{X}} = \sqrt{a_k^2 + b_k^2} \text{ and } \mathrm{A}_k^{\mathrm{Y}} = \sqrt{c_k^2 + d_k^2} \quad (7)$$

where *ab* and *cd* are the coefficient pairs for X and Y, respectively.

In this case, no assumption is made concerning the relationship between the two signals. This relationship is expected to be brought out in the correlation between two pairs of amplitudes.

The second approach previously used is to treat X and Y as independent and to combine the four coefficients linearly:

$$\mathrm{A_k} = \sqrt{a_k^2 + b_k^2 + c_k^2 + d_k^2} \quad (8)$$

However, this hypothesis of two independent signals is unsatisfactory because both signals relate to the same distance along the outline. We therefore prefer not to use this technique. The third descriptor (EFA A) is a combination of 2K amplitudes (Eqn 7). This EFA was analysed using NTSYS-pc (Rohlf, 1993).

Complex discrete Fourier analysis

This method has seldom been used since its development in morphometrics (e.g. Kincaid & Schneider, 1983; Bertin *et al*., 2002) although it offers the great advantage of processing a two-dimensional signal in a single pass. This analysis was performed using MATLAB Toolbox CDFT 2.7 (Dommergues, 2001) with modifications made to the code. Unlike harmonic decomposition in the realm of real numbers (e.g. elliptic analysis), harmonics derived from a complex signal do not have conjugate symmetry. So N conjugate harmonics (or N/2 pairs of conjugate harmonics) are obtained. Two pairs of coefficients are obtained for each pair of conjugate harmonics (Eqn 5). Thus, the descriptor obtained (CDFA C) corresponds to the combination of 4K coefficients (2*2K, where K = number of conjugate pairs of harmonics). The amplitude is obtained from the modulus of the Fourier coefficient (Eqn 6). The final descriptor is obtained by combining 2K amplitudes.

There is a correspondence between the elliptic coefficients and coefficients from the complex method that can be obtained by comparing the linearity of *F*:

$$F(\mathrm{Z}) = F(\mathrm{X}) + iF(\mathrm{Y}) \quad (9)$$

with

$$F(Z) = \sum_{\mathrm{k}=0}^{\mathrm{N}-1}\left(\frac{1}{\mathrm{N}}\sum_{j=0}^{N-1} Z_j e^{-i\frac{2\pi kj}{N}}\right)e^k \quad (10)$$

$$F(X) = \sum_{k=0}^{N-1}\left(\frac{1}{N}\sum_{j=0}^{N-1} X_j e^{-i\frac{2\pi kj}{N}}\right)e^k \quad (11)$$

$$F(Y) = \sum_{k=0}^{N-1}\left(\frac{1}{N}\sum_{j=0}^{N-1} Y_j e^{-i\frac{2\pi kj}{N}}\right)e^k \quad (12)$$

Having identified the base e^k, we have:

$$C_k(Z) = C_k(X) + iC_k(Y) \quad (13)$$

If we write these coefficients in their developed form:

$$C_k(Z) = A_k + iB_k \quad (14)$$

$$C_k(X) = a_k + ib_k \quad (15)$$

$$C_k(Y) = c_k + id_k \quad (16)$$

that is to say:

$$A_k + iB_k = (a_k + ib_k) + i(c_k + id_k) \quad (17)$$

$$= (a_k + c_k) + i(b_k + d_k) \quad (18)$$

Thus if we obtain the modulus of Equation 14, we recover the general formulation of the amplitude (Eqn 6). Substituting the terms with those of Equation 18 gives the following correspondence:

$$\sqrt{A_k^2 + B_k^2} = \sqrt{(a_k + c_k)^2 + (b_k + d_k)^2} \quad (19)$$

$$= \sqrt{a_k^2 + b_k^2 + c_k^2 + d_k^2 - 2(a_k d_k - b_k c_k)} \quad (20)$$

In this form (Eqn 20), we recover the second form of the amplitude from the elliptical method, although in this instance a supplementary term corresponding to the connection between the two real signals is introduced. Elliptical coefficients can therefore easily be transformed into coefficients of complex methods, if results have not been truncated to N/2. For all the outline descriptors, a preliminary truncation is performed and ten harmonics are used.

Finally, following the definition of Lestrel (1997b) but with a restricted number of attributes [Form = (Size, Shape)], five descriptors of form (Traditional, CDFA A nstd, CDFA C nstd, EFA A nstd and EFA C nstd) and five descriptors of shape (Log Shape Ratio, CDFA A std, CDFA C std, EFA A std and EFA C std) have been defined and used in this study.

METHODOLOGICAL FRAMEWORKS

STATISTICAL CLASSIFICATION

Linear DA is the method most frequently used in species discrimination studies (e.g. Carrasco, 2000) as well as in comparative studies of morphometric descriptors (McLellan & Endler, 1998; Dommergues *et al*., 2003). In these approaches, evaluation of taxa or comparison of descriptors is based on the quality of the discriminant model. Frequently, this quality is based on the percentage of correct classification (or incorrect classification for error rate) using a resubstitution method. However, resubstitution overestimates quality because the model is not independent of classified individuals and bias increases with the number of parameters incorporated in the model (Lebart, Morineau & Piron, 2000). For example, Lance, Kennedy & Leberg (2000) have shown that this bias may have a strong influence on taxonomic studies. Likewise, and as emphasized by the difference in the number of parameters, such bias may have a strong effect on the evaluation of various morphometric descriptors.

The aim of this section is to quantify the expected predictive power, or conversely the expected error rate, of a descriptor when statistical classification is used in fossils of present-day species or in recent unidentified samples extracted from owl pellets. If we are to choose the more accurate descriptor, such bias must be corrected. This is done by the leave-one-out cross-validation technique. This procedure has been preferred to the sample splitting techniques in training and test samples because, unlike splitting techniques, it does not alter the training model (Lance *et al*., 2000). For each sample (N–1), a linear DA is performed and the class of the deleted observation evaluated using a posteriori probabilities of 'plug-in' classifiers (Ripley, 1996). As there are no grounds for considering that all models are linear, quadratic models are evaluated in the same way.

Bias associated with a descriptor is calculated using the procedure outlined by Lance *et al*. (2000). Random permutations between species are made and the complete analysis is repeated. In this way, the expected error rate is 50% (Lance *et al*., 2000) and can be compared with resubstitution and leave-one-out cross-validation results. Bias is taken to be the deviation between the expected 50% value and the actual value. Analyses are performed under MATLAB using the Discriminant Analysis Toolbox of Kiefte (1999).

MORPHOSPACES AND PATTERNS INFERRED

Morphospace analysis is now used increasingly in palaeobiology; see, for example, the many disparity-based studies in macroevolution. In many cases, these morphospaces are constructed from a whole family of morphometric descriptors. McGhee (1999) defined them as empirical morphospaces and roundly criticized their utilization in the study of evolutionary patterns (but see Eble, 2000b for a review).

Dimensionality of morphospace

The main criticism of the use of empirical morphospaces is their dependence on samples, entailing non-comparability and instability. This dependence makes it a prerequisite to reduce dimensionality of

morphospace before it can be analysed. This involves considering initial dimensions as essentially supporting signal with a minor effect of noise; suppressed dimensions essentially reflect noise. In this way an optimal pattern can be analysed free of noise (on which sample dependence has a strong effect).

In many cases, the choice of dimensionality is viewed as subjective. However, some methods have been genuinely effective (Jackson, 1993). One method relates to detecting structured as opposed to random components (broken-stick model). Others relate to the stability of components in the face of perturbation of the sample using non-parametric bootstrapping (Efron & Tibshirani, 1993). In these cases, PCA was repeated 1000 times. The methodology outlined by Jackson (1993) is first to retain a component with an eigenvalue different from that of its successor (i.e. ranges of two successive eigenvalues should not overlap). Second, components with two or more stable eigenvector coefficients are chosen (i.e. 95% confidence limits that do not contain zero).

Data perturbation may involve axis reflection as well as permutations and rotations between axes with similar eigenvalues. Axis reflection is not a true perturbation and must be corrected for. The simplest method is to fix the sign of the largest coefficient (absolute value) in initial space (Mehlman, Shepherd & Kelt, 1995). This method, employed here, takes rotations and permutations to be true perturbations contrary to other suggested methods such as Procrustes analysis or partial bootstrap analysis (Lebart *et al*., 2000). These other methods either adjust homologous axes or conserve initial space and thus minimize perturbations due to rotation or permutation.

All three methods are used here to extract dimensionality from the various morphospaces derived from PCA. Analyses are performed on the correlation matrix. Using the correlation instead of the variance/covariance matrix suppresses information about the amount of variance and gives equal weight to all parameters in the analysis. It implies that tiny aspects of form have a similar influence to that of other aspects with larger dispersions, while tiny aspects of shape can be more informative than large ones.

With all parameters in the same units, and use of the variance/covariance matrix appropriate in context, we use this approach because with linear descriptors differences between variances can be related to the definition of measurements. It also limits the effect of size. All the procedures are implemented using MATLAB code. Variances of principal components have not been rescaled to unity and so are equal to their eigenvalue. In this way, components do not have the same weight in subsequent analyses and this weight corresponds to their initial degree of variance.

Similarity of morphospaces

Because morphospaces are based on descriptions of the same morphological structure, they should be very similar. Dissimilarity may result in either an enhanced representation of the morphological structure or an improved recombination of information by one descriptor or another. Similarity can be observed on the basis of covariation between morphospaces. One approach is to perform a partial least squares (PLS) analysis (Sampson *et al*., 1989; McIntosh *et al*., 1996; Rohlf & Corti, 2000). This approach, which was developed to explore patterns of covariation between two or more sets of variables, has been widely used in neurobiology (e.g. McIntosh *et al*., 1996; McIntosh & Gonzales-Lima, 1998; Della-Maggiore *et al*., 2000; Lobaugh *et al*., 2001; Nestor *et al*., 2002). In morphometrics it has been used in several studies to investigate covariation between shape and trophic variables (Adams & Rohlf, 2000; Rohlf & Corti, 2000), shape and environmental variables (Corti *et al*., 1996), or between two blocks of shape variables such as skull and limb bones (Rohlf & Corti, 2000), dorsal and ventral views of skulls (Rohlf & Corti, 2000), or developmental modules of fly wings for a quantified level of integration (Klingenberg & Zaklan, 2000; Klingenberg *et al*., 2001).

Two-block PLS analysis corresponds to the singular value decomposition (SVD) of the cross-covariance (or cross-correlation) matrix between two sets of variables. The extended computational procedure is given in the literature; here, we simply underline the major features. SVD provides sets of mutually orthogonal latent variable (LV) pairs, that account for progressively less of the sum of the squared cross-block covariance (McIntosh & Gonzales-Lima, 1998). The number of LVs is fixed by the minimal dimensionality of morphospaces compared and goodness of fit can be measured by the ratio between the sum of the squared cross-block covariance accounted for by k LVs (corresponding to the sum of k^2 singular values) and the total squared cross-covariance (corresponding to the sum of total squared singular values). The value of each variable of one block (called salience) corresponds to its correlation with the latent variable of the other block (Tabachnick & Bookstein, 1990).

As in PCA, scores on LV pairs can be computed and contrasted. The scores of each block readily permit the computation of correlations between the paired variables. The significance of latent variables and correlations between blocks can be assessed by a sampled randomization test (Sokal & Rohlf, 1995) using 5000 permutations (Rohlf & Corti, 2000). The reliability of the contribution of saliences is assessed using the ratio of salience to bootstrap standard error (1000 resampling with replacement) as outlined by Lobaugh

et al. (2001). In our study, analysis was performed using MATLAB based in part on functions written by A. R. McIntosh.

Two-block PLS has already been used to quantify relationships between morphospaces (Tabachnick & Bookstein, 1990). Those authors observed different aspects of shape and considered the initial morphometric descriptors to be shape variables. Here, we use PLS for the same quantified aspect of shape or form using different descriptors. The input data are not the initial morphometric descriptors but the scores on p and q stable PCs defining the two morphospaces to be compared. This technique has been chosen because our purpose is not to explain one morphospace in terms of another but to define whether the different morphospaces contain similar information about the distribution of specimens considering all potential sources of dissimilarity (extraction of information, mathematical processing, behaviour in the face of constraints of multivariate ordination).

Congruence of pattern inferred

Morphospaces provide the basis of disparity analysis and their bias can modify observed disparity patterns. The disparity framework features a number of problematic and hierarchical levels of study, e.g. evolutionary radiation (Foote, 1999), extinction selectivity (Foote, 1991a, 1992), relative importance of development in evolution, and of organization in dynamics (Eble, 2002). However, in these fields, problems of pattern stability with regard to anatomical structure, hierarchical level and temporal resolution changes can be outlined and have sometimes been tested (Foote, 1994, 1995, 1999; Eble, 2000a).

Another question is raised, concerning how robust patterns may be when descriptors within the same structure are changed. In this case, modification of pattern cannot be inferred from the possibility of disjoint evolution between the different morphological structures showing different evolutionary constraints. It must be inferred from the variation in the quantity and structure of morphological information contained in the descriptors and in their recombination. Morphospace/disparity analyses can easily be applied to other problems and in such cases this question remains. Thus, although our study is not concerned with disparity, the problem is similar and can be viewed in terms of whether or not the relationship between intraspecific dispersions within the global morphospace is affected by the descriptor used.

To explore this issue, we employed the classic parameter of dissimilarity used in disparity studies (Foote, 1991a; Eble, 2000a): the sum of variances or total variance (i.e. the sum of the variance of each component). In the initial framework of the disparity study (i.e. disparity changes through time), various tests on disparity values have been used. Some are non-parametric (Foote, 1990, 1991b, 1993; 1994; Wills, Briggs & Fortey, 1994; Eble, 2000a). A parametric approach based on the z-test has been also used (Foote, 1993; Eble, 2000a, 2002).

Here, the test of pattern robustness is based on the differences in dispersion between the two species observed for each descriptor. Two ways of testing difference are used. First, a sampled randomization test (Sokal & Rohlf, 1995) is used to test a simple observed difference in total variance (TV). The observed simple differences between the two species are compared with the 5000 values obtained by randomization. The second technique is more akin to procedures previously used in disparity studies (Foote, 1993; Eble, 2000a, 2002). Bootstrap estimators of TV are computed (with 200 bootstrap replicates) and z-tests are performed. This kind of parametric approach is based on the fact that bootstrap-sample statistics are normally distributed even if the original distributions are non-normal and so the z-statistic follows a centred normal distribution (Efron & Tibshirani, 1993). However, it seems possible that highly multimodal morphological distribution (as should be observed in higher-level studies) may bias a normal approximation of the bootstrap statistic and thus invalidate the hypothesis of normality. This implies that the z-statistic does not necessarily follow the centred normal distribution.

In the disparity framework nothing indicates that species in the groups under study (stratigraphic or otherwise) do not have multimodal distributions in P-dimensional morphospace; thus the bootstrap sample of TV follows a normal approximation. Because the aim is to explore the effects of descriptor changes in the same way as can be done at various levels of morphospace study (i.e. from macroevolutionary to microevolutionary), this assumption is maintained here. Distribution of the z-statistic for the bootstrap estimator of TV is considered to be unknown and so an approach by sampled randomization is adopted. The observed z-statistic is compared with 5000 values obtained by sample randomization and, for comparison, compared with critical values obtained from centred normal approximation.

All tests are two-tailed, with multiple comparisons performed on the same specimens. Descriptors represent, in part at least, different codings of the same initial biological information. Thus, each descriptor has some information in common with the others; the morphological information processed is highly non-independent and probabilities probably reflect a relatively high risk of Type I error.

RESULTS

STATISTICAL CLASSES

Apparent rate and bias estimation

All estimates of apparent quality (Wilks' Λ, Mahalanobis distance and rate of classification using resubstitution) of the discriminants (Table 1) are good and homogeneous among descriptors. The non-standardized Fourier coefficients based on the complex method (CDFA C nstd) yield the best results. Generally, the standardized amplitudes based on the elliptical method (EFA A std) give the poorest results in each case, except for those based on resubstitution of the linear discriminant where its version based on the complex method (CDFA A std) is poor. In each case, the non-standardized version provides a better result than the standardized one. Linear descriptors yield relatively good results mid-way between amplitude-based and coefficient-based descriptors. Generally, CDFA gives better results than EFA in terms of relative descriptors.

Previous results are based on the apparent rate, which is biased by the number of parameters and the non-independence of observations (see Methodological frameworks). To estimate this bias, a randomization procedure was performed. The expected value of correct classification is 50%. The resubstitution technique overestimates this rate (Table 2); the rate of correct classification yields a similar value for descriptors with the same number of parameters and this value is positively correlated with this number. The bias (viewed as the difference between expected and actual values) is larger with the quadratic discriminant where the observed rate is close to the results obtained previously without randomization. Values of correct classification using the leave-one-out cross-validation techniques are close to the expected value of 50%. The bias in this case is generally less than 1% and only the larger descriptors (Fourier coefficients) display a greater bias with the quadratic discriminant (close to 3%). Moreover, standard deviations of the results are larger in this case (5% on average) than with the linear discriminant (3% on average).

The number of parameters largely biases the resubstitution method. It is inadequate for estimating the model's quality and for purposes of comparison. In contrast, leave-one-out cross-validation gives relatively unbiased estimates and so allows effective comparison with various morphometric or statistical methods (i.e. linear vs. quadratic discriminants).

Linear vs. quadratic discriminants

For each descriptor, the results of cross-validation on linear discriminants are better (Table 1) than for quadratic discriminants (from 0.49% for CDFA A nstd to 9.68% for CDFA C nstd, with an average of 4.02%). Moreover, linear discriminants seem to be more stable than quadratic ones, as the standard deviations obtained from randomization indicate. Thus, it would be better to observe differences between descriptors on linear discriminants and for all further comparisons to be made on cross-validation estimates based on linear discriminants.

Standardized vs. non-standardized versions

In non-standardized versions, CDFA C yields the better value (98.99%) and CDFA A the poorer one (96.02%). In standardized versions, the better descriptor is CDFA C (96.39%) and the poorer one is EFA A (89.05%). Nonstandardized versions of each descriptor give a better rate of correct classification than the

Table 1. Results of the linear and quadratic discriminants. NP, number of parameters incorporated in the model. Rsub, results of the resubstitution method. Cval, results of the leave-one-out cross-validation method. Rsub and Cval show the percentage of correct assignment using a posteriori probabilities obtained from linear and quadratic discriminant. D^2, Mahalanobis distance between the two species. All values of Wilks' Λ are highly significant at $P < 0.0001$

		Linear			Quadratic		
Descriptor	NP	Wilks' Λ	D^2	Rsub	Cval	Rsub	Cval
Traditional	23	0.119	29.18	99.78	98.35	99.91	95.29
Log Ratio	23	0.154	21.58	98.78	96.20	99.36	93.45
CDFA A nstd	20	0.154	21.67	98.84	96.02	99.96	95.53
CDFA A std	20	0.244	12.19	94.03	90.53	98.31	89.11
CDFA C nstd	40	0.072	50.69	100	98.99	100	89.31
CDFA C std	40	0.117	29.54	99.90	96.39	100	89.66
EFA A nstd	20	0.186	17.24	98.96	96.22	99.77	93.64
EFA A std	20	0.262	11.08	94.35	89.05	98.05	86.73
EFA C nstd	40	0.112	31.18	99.93	97.36	100	89.52
EFA C std	40	0.123	28.02	99.89	95.09	100	91.68

standardized ones (Table 1). The difference ranges between 2.15% (Traditional vs. Log Shape Ratio) and 7.17% (EFA A). Amplitude-based descriptors (CDFA A and EFA A) have larger differences (respectively 5.49% and 7.17%) as opposed to *c.* 2% for coefficient-based and linear-based descriptors. The loss in quality due to standardization is easily explained by the fact that a large difference size and shape between the two species is due to significant difference in size (ANOVA: $F\sqrt{\sum V^2} = 147.23$ $P < 0.0001$ and $F\sqrt{\sum V^2} = 168.29$ $P < 0.0001$) with *M. agrestis* larger than *M. arvalis*. The largest differences observed for the amplitude-based descriptor appear to indicate that after removing the size effect, these descriptors no longer displayed sufficient aspects of shape for the discrimination of species [which seem to be present in other descriptors such as the coefficient-based or Log Shape Ratio (LSR) descriptors].

To summarize the results a UPGMA tree was constructed from the absolute difference in the percentage of correct classification due to cross-validation (Fig. 2). Major structuring of the tree reflects size effects. However, the impact of other factors appears to be significant since the tree was not symmetrical owing largely to the poor results obtained by amplitude-based descriptors. Amplitude is a technique that totally removes the effects of orientation and quantifies only the magnitude of the harmonics. Thus, amplitude-based descriptors eliminate many aspects of shape; for example, no information about symmetry/asymmetry is conserved. In contrast, coefficient-based descriptors maintain this information, which is also present in linear descriptors. This information seems to be the second major effect that structured the tree and explained the similarity in results for linear and coefficient-based descriptors.

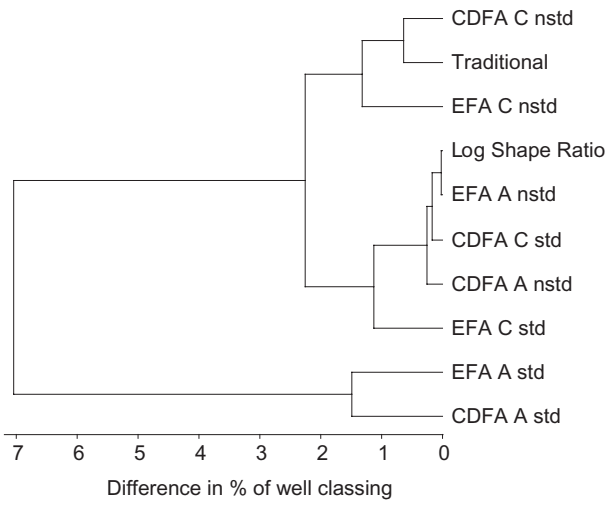

Figure 2. UPGMA tree based on absolute differences in the percentage of correct classification (Leave-one-out cross-validation) obtained with a linear discriminant. Major effects are the size factor (standardized vs. non-standardized descriptors) and tooth symmetry/asymmetry (amplitude-based vs. coefficient-based and linear-based descriptors). There appears to be a slight effect due to the Fourier method used (elliptic vs. complex), but none due to the type of morphological information extracted (outline vs. linear).

Table 2. Bias in expected results of the linear and quadratic discriminants. For each randomization (1000), discriminant and classification by resubstitution or leave-one-out cross-validation was performed for each descriptor. Because observations were randomly re-attributed to one of the two species, the expected rate of correct classification is 50%. Bias can be viewed as the deviation between the actual and the expected rates. NP, number of parameters incorporated into the model. Rsub, results of the resubstitution method. Cval, results of the leave-one-out cross-validation method. Rsub and Cval show the average percentage of correct attributions using a posteriori probabilities obtained from linear and quadratic discriminants. Numbers in brackets indicate the standard deviation

		Linear		Quadratic	
Descriptor	NP	Rsub	Cval	Rsub	Cval
Traditional	23	59.46 [2.48]	49.55 [2.88]	90.80 [1.93]	49.06 [5.08]
Log Ratio	23	59.41 [2.52]	49.48 [2.91]	90.66 [1.99]	49.00 [4.82]
CDFA A nstd	20	58.34 [2.46]	49.70 [2.81]	87.39 [2.10]	49.27 [5.00]
CDFA A std	20	58.34 [2.43]	49.72 [2.78]	86.69 [2.12]	49.19 [4.83]
CDFA C nstd	40	67.10 [3.35]	49.53 [3.97]	99.61 [0.43]	47.40 [5.88]
CDFA C std	40	67.11 [3.39]	49.53 [4.01]	99.65 [0.40]	47.21 [5.76]
EFA A nstd	20	58.33 [2.41]	49.68 [2.75]	87.94 [1.97]	49.35 [4.94]
EFA A std	20	58.32 [2.38]	49.68 [2.72]	87.61 [2.07]	49.26 [4.88]
EFA C nstd	40	67.13 [3.32]	49.53 [3.95]	99.63 [0.41]	47.35 [5.94]
EFA C std	40	67.14 [3.35]	49.57 [4.01]	99.71 [0.37]	47.50 [6.00]

CDFA vs. EFA

CDFA and EFA yield similar results for their respective versions of descriptors (Table 1). In most cases, slight differences favour the complex method (average of 1.47% without A nstd). Only the non-standardized amplitude method failed to reveal this difference (96.02% for CDFA and 96.22% for EFA). This gain in favour of CDFA is recovered by the structuring of the more internal branch of the UPGMA tree (Fig. 2).

Linear vs. Fourier analysis

Although CDFA C has slightly better predictive power than linear-based descriptors (98.99% against 98.35% for nstd and 96.39% against 96.20% for std), the differences are small. So, no effect appears to be due to the type of morphological information extracted.

Morphospaces and patterns inferred

Dimensionality of morphospaces

The various methods employed to reduce the dimensionality of morphospace (broken-stick, bootstrap eigenvalues and eigenvectors) yield results in the same order from two to five axes (Table 3). For each descriptor, the methods show a difference from one to three axes (with a majority in one axis). These results differ from those obtained using thresholds of 75% (4–7 axes) or 95% (11–17 axes) as might be expected from results obtained by Jackson (1993), which underlined that these threshold methods did not estimate dimensionality correctly.

Table 3. Dimensionality of morphospaces from various threshold methods. Results obtained from broken-stick (BS), bootstrap of eigenvalues (Eva), and eigenvector (Eve) coefficient methods are similar. These three methods are considered as the most consistent by Jackson (1993). Thresholds of 75% and 95% of total variance are given for comparison. N retained corresponds to the dimensionality of reduced morphospace considered in subsequent analyses

				Bootstrap	
Descriptor	75%	95%	BS	EVa	EVe
Traditional	5	12	2	2	3
Log Ratio	7	14	4	1	4
CDFA A nstd	5	11	3	3	4
CDFA A std	5	12	2	2	3
CDFA C nstd	5	15	4	2	5
CDFA C std	6	16	3	2	3
EFA A nstd	4	12	3	4	5
EFA A std	5	13	2	2	3
EFA C nstd	5	16	4	2	5
EFA C std	5	17	3	2	4

In the majority of cases where Fourier-based descriptors and broken-stick or bootstrap eigenvector methods are used, morphospaces lose one axis between the non-standardized and standardized versions (except for EFA A and CDFA C using the eigenvector method, where two axes are lost). Morphospaces based on linear descriptors show increased dimensionality between non-standardized and standardized versions; reduced dimensionality can be related to the loss of the size factor. Conversely, the gain in dimensionality of the linear descriptor may be related to bias in the recombination of information by PCA. The considerable redundancy of non-standardized linear variables may affect structuring or stability and focus information on the first axis. Standardization, in this case, redistributes information and a better representation is obtained by PCA. For each descriptor, the bootstrap eigenvector method generally gives one more axis than the broken-stick method. Frequently, this axis features fewer stable variables than for higher-ranking components. Moreover, these stable eigenvector coefficients have similar values. It seems that these coefficients represent redundant morphological information and the dimensionality may be overestimated.

The eigenvalue method yields results with loss of one (linear and CDFA A) or two axes (EFA A). For other descriptors no loss of axis is observed; two axes are retained in most cases. This method underestimates dimensionality compared with the other two. This can be explained by the fact that although dimensions are structured and stable, eigenvalues are similar and so ranges overlap. The cut-off point of the dimensionality is a compromise between components which mostly have signal and those which mostly consist of noise. Broken-stick is a simple procedure based on the recognition of structured components, in opposition to the expected distribution of eigenvalues from a random signal. However, it does not take sample size into account and thus gives results between those of the two other procedures based on sampling error. We prefer to use the broken-stick procedure, which gives structured axes that are stable (lower estimates than eigenvectors) but may present similar eigenvalues (higher estimates than eigenvalues).

Complex and elliptic Fourier descriptors present equal dimensionality of their morphospace. However, a reduction in dimensionality can be observed between coefficient-based and amplitude-based descriptors. This may be due to the loss of information when the latter are used, i.e. structured in a less complex way (fewer variables). The case of linear measurements (Traditional and LSR) seems to indicate that the structuring of data affects the capacity of PCA to recombine information (problem of the size factor).

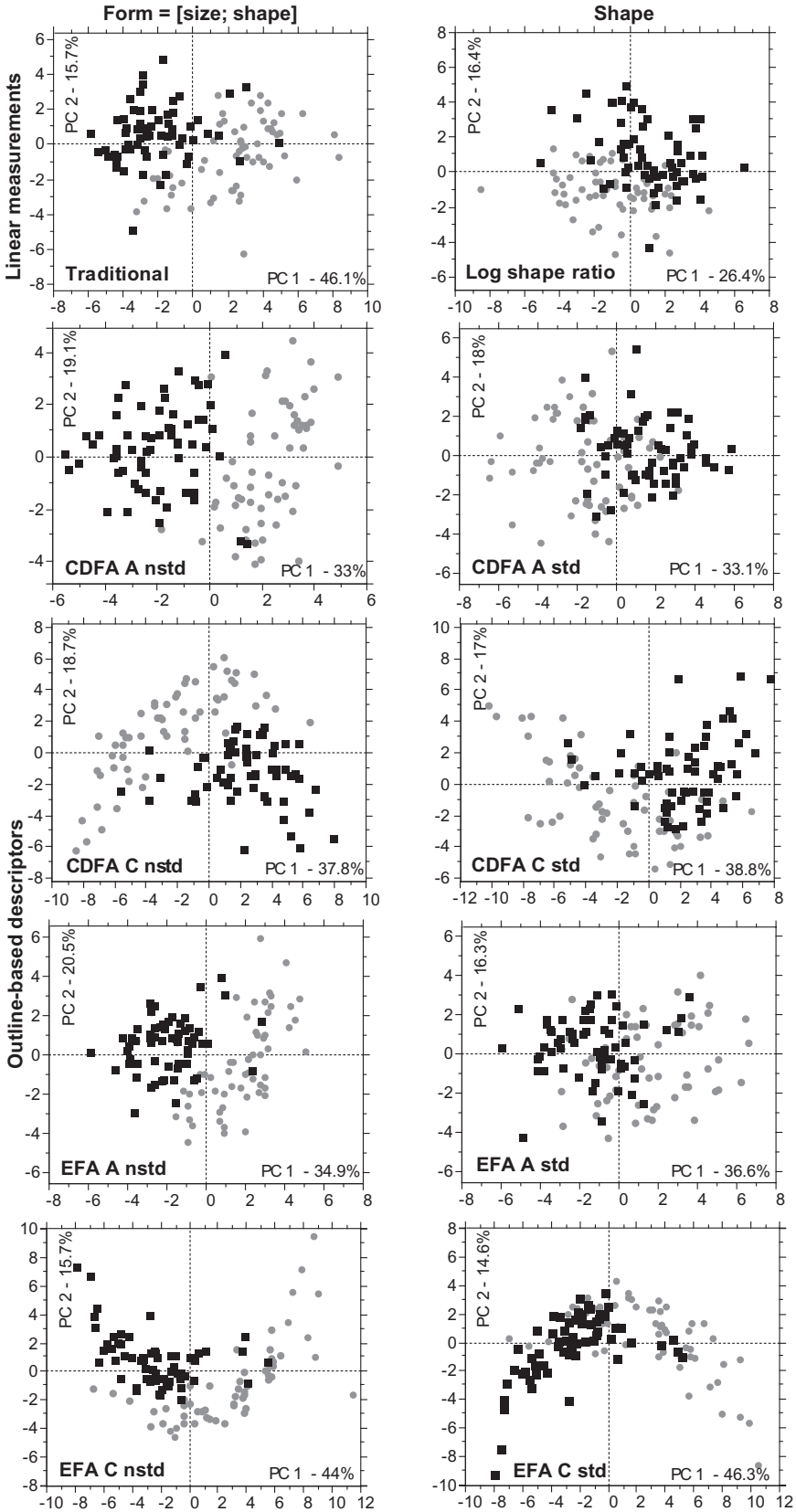

Figure 3. Morphospace of *Microtus agrestis* (●) and *M. arvalis* (■); the first two PCs from each descriptor.

Table 4. Percentage of the initial variance summarized by each morphospace. D, dimensionality

Descriptor	D	Per cent initial variance
Traditional	2	61.80
Log Ratio	4	62.81
CDFA A nstd	3	67.58
CDFA A std	2	51.19
CDFA C nstd	4	72.20
CDFA C std	3	64.04
EFA A nstd	3	69.25
EFA A std	2	52.88
EFA C nstd	4	74.04
EFA C std	3	68.30

Similarity of morphospace

The first two component spaces (i.e. PCs 1 and 2) accounted for similar percentages of the initial variance among descriptors (Fig. 3). PC 1 represents from 26.4% (LSR) to 46.3% (EFA C std) of the total variance and PC 2 from 14.6% (EFA C std) to 20.5% (EFA A nstd). For outline-based descriptors, the percentage of the total variance accounted for by PC 1 increases (1–2%) between non-standardized and standardized versions. The percentage decreases for PC 2 (1–4%). The reverse is observed for linear-based descriptors (decrease of 20% for PC 1 and increase of 0.7% between non-standardized and standardized versions). Complete morphospaces (i.e. all dimensions retained) accounted for between 51.18% (CDFA A std) and 74.05% (EFA C nstd) of the initial variance (Table 4); the variation in this percentage is largely due to the difference in dimensionality.

However, the two complete linear-based morphospaces (PCs retained) show a similar percentage of the total variance (61.8% for Traditional and 62.81% for LSR) despite the difference in dimensionality. This, together with the previous difference between outline and linear-based descriptors, appears to indicate that there is considerable redundancy of linear measurements when these variables are non-standardized, affecting recombination of information by PCA. Information is condensed in PC 1 due to size effects and PCA fails to recover stable, independent morphological components. When standardization occurs, relationships between variables are better represented and better segregated among components.

The various morphological mappings constructed display a similar general pattern. In all cases, much of the interspecies variance is based on PC 1, along which both species have virtually separate distribution. Observed distinctness (Fig. 3) seems larger when non-standardized descriptors are used rather than standardized ones; this is in accordance with the difference in size reported previously. Thus, much of the total variance observed in the initial matrices is interspecies variance.

In most cases (39/45 pairwise comparisons), PLS did not reduce dimensionality (i.e. all LVs are significant at $P < 0.05$, and in many cases at $P < 0.0001$). When reduction occurs, a loss of 1–2 pairs of latent variables from the maximum possible is observed. In four cases of reduced dimensionality, the LSR-based morphospace is considered in comparison with the other block as CDFA A nstd, CDFA C nstd and std, EFA A nstd. The other two cases are comparisons between CDFA C std and either CDFA A nstd or EFA A nstd. The pair of latent variables accounts for an average of 85.86% (first pair) and 98.52% (first two pairs) of the total covariation between the two morphospaces. Thus, LV 1 accounted for most of the covariation between morphospaces, although LV 2 accounted for a non-negligible part (maximum 34%), even when there are two LVs. Scatterplots of the first two pairs of LV (Fig. 4) show that LV 1 summarizes the cross-covariance between two morphospaces due to interspecific differentiation, while LV 2 summarizes another aspect of the cross-covariance due to intraspecific differentiation.

In many cases, the bootstrap ratio of saliences shows a major contribution to one PC, but this evidence of similarity is frequently altered (i.e. contribution of more than one PC) when some initial factors of dissimilarity have been introduced (i.e. linear vs. Fourier; std vs. nstd, coefficients vs. amplitude). The one-to-one correspondence between PCs of each block seems less affected by the mathematical processing of outline (elliptic vs. complex). However, the large number of comparisons makes this information difficult to summarize and it is not developed further here.

A more obvious approach is to observe the correlation coefficient between blocks on the first two LVs (Table 5) as a proxy for similarity. On both LVs, greater similarities (i.e. high correlation coefficient) are obtained from comparison within outline-based descriptors than from between linear- and outline-based descriptors. On LV 1, correlations are relatively high even between linear and outline morphospace (from 0.56 for Traditional vs. EFA C std to 0.86 for Traditional vs. CDFA A nstd). Strangely, the correlation between standardized linear measurements morphospace (LSR) and outline-based morphospace is higher with non-standardized versions of outline-based descriptors than with standardized ones.

Correlations decrease largely on LV 2 (from 0.22 for Traditional vs. CDFA A std to 0.48 for Traditional vs. CDFA A nstd). Thus, between morphometric families (i.e. linear measurements vs. outline), morphospaces present relatively good similarity, although in only one

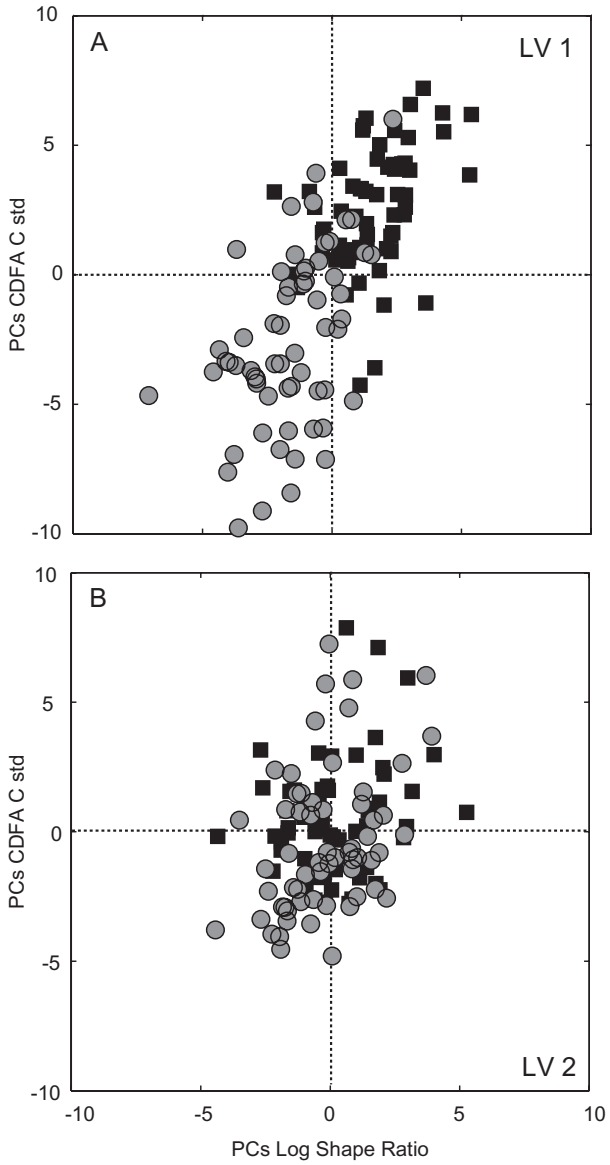

Figure 4. Example of scatterplot of the two first pairs of latent variable. ●: *Microtus agrestis*. ■: *M. arvalis*. A, LV 1 B, LV 2. Ordinates correspond to CDFA C std-based morphospace and abscissa to Log Shape Ratio-based morphospace. The correlations are 0.7 (A) and 0.3 (B). Each scatterplot corresponds to one aspect of covariation between morphospaces: interspecific differentiation on LV 1 (A) and intraspecific variability on LV 2 (B).

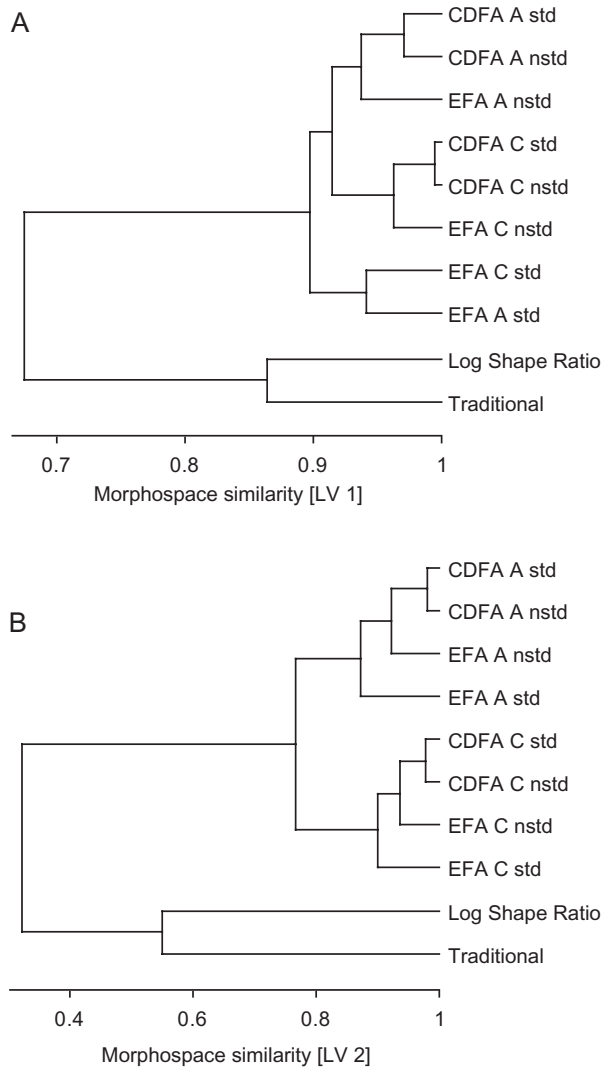

Figure 5. UPGMA trees based on morphospace similarity. Similarity is considered as the correlation between each morphospace (i.e. each block) obtained from two-blocks partial least squares analysis. Similarity is based on the results of the first latent variables (A) or the second latent variables (B). In both cases, the information extracted (i.e. linear measurements vs. outline) is the more important feature structuring the trees.

respect, that of interspecific differentiation. In contrast, within outline-based morphospaces, similarities are very great. From 0.86 for EFA A std vs. CDFA C nstd to 0.99 for the two versions of CDFA C on LV 1, and from 0.62 for CDFA A nstd vs. EFA C std to 0.98 for the two versions of CDFA C on LV 2.

To summarize all pairwise similarities, a UPGMA tree was constructed for each pair of LVs (Fig. 5). The major effect of dissimilarity (albeit relatively weak since correlations are close to 0.7) structuring the two trees corresponds to the type of information extracted (i.e. linear measurements vs. outline). Dissimilarity due to size standardization is well marked in linear-based descriptors and in elliptic-based descriptors. Complex-based descriptors are affected very little by standardization. Another factor structuring trees is the dichotomy between amplitude-based and coefficient-based descriptors. Thus, the inclusion of orientation (tooth symmetry/asymmetry) in outline-

Table 5. Similarities between morphospaces. Similarity is viewed as the correlation coefficient between blocks on each latent variable (LV). The bottom triangular matrix corresponds to LV 1 and the top one to LV 2. All LV 1 and associated correlations are significant at $P < 0.0001$. All LV 2 and associated correlations are significant at $P < 0.0001$ for comparison among outline-based descriptors: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$. The first line under the correlation corresponds to the probabilities of the latent variables and the second line corresponds to the probabilities of the associated correlations. Probabilities correspond to the number of permutations out of 5000 exceeding the observed value. Trad, traditional; LSR, Log Shape Ratio

			CDFA				EFA			
Log	Trad	LSR	A nstd	A std	C nstd	C std	A nstd	A std	C nstd	C std
Trad		0.5486	0.4813	0.2225	0.4377	0.2494	0.4860	0.2311	0.4844	0.1485
****	****	****	****	***	****	***	****	*		
****	****	****	****	**	****	***	****	0.089		
LSR	0.8637		0.2371	0.2531	0.3681	0.3165	0.2268	0.2547	0.3553	0.4165
0.08	****	**	****	0.06	*	*	****			
*	****	***	****	*	****	***	****			
CDFA A nstd	0.8576	0.7512		0.9814	0.8320	0.7988	0.9135	0.8508	0.7968	0.6244
CDFA A std	0.4686	0.6422	0.9705		0.7930	0.7849	0.9305	0.8660	0.7345	0.6487
CDFA C nstd	0.7611	0.7599	0.9272	0.9056		0.9765	0.8206	0.7625	0.9430	0.8677
CDFA C std	0.5771	0.7016	0.9048	0.9163	0.9946		0.8026	0.7586	0.9313	0.8989
EFA A nstd	0.8224	0.7234	0.9418	0.9327	0.9139	0.9115		0.9024	0.9088	0.6865
EFA A std	0.4609	0.6026	0.9021	0.9347	0.8586	0.8703	0.9345		0.742	0.7798
EFA C nstd	0.6933	0.7317	0.9034	0.8971	0.9656	0.9609	0.9499	0.8794		0.9315
EFA C std	0.5611	0.688	0.8624	0.8791	0.9103	0.9160	0.8930	0.9415	0.9336	

based descriptors affects the degree of similarity among morphospaces.

Congruence of pattern inferred

The raw values of difference in intraspecific variance (i.e. difference in the sum of variances for each component) cannot be directly compared among morphospaces because it did not take the same number of axes into account. Moreover, axes were not rescaled to unity; the variance of the axis thus represented the initial part, which is dependent on the number of initial variables. However, as a first approach, the signs of the difference may be compared. In many cases (except LSR), differences are positive: i.e. *M. agrestis* displays greater intraspecific variance than *M. arvalis* (Table 6). On the other hand, patterns of significance of the difference can be compared. Similar results are obtained whichever test is used (randomization of simple difference, randomization or normal approximation of *z*-statistic). However, the test based on randomization of simple differences seems more conservative than the other two based on the *z*-statistic. The two approaches to assessing the significance of the *z*-statistic yield similar results, although slightly more conservative values are obtained from randomization. It should be emphasized that the standard assumption for the *z*-statistic seems to be correct in our case.

Differences from non-standardized descriptors are always significant. In the majority of cases, standardized descriptors show no significant differences in intraspecific variance, except for amplitude-based descriptors (EFA A and CDFA A). Thus, generally (except for amplitude-based descriptors), when difference in size is removed, the intraspecific variances in shape of the two species are finally equivalent.

DISCUSSION

Our analyses show that effects of descriptor changes are diverse and depend on many initial factors of dissimilarity between them, e.g. information extracted, size standardization, initial mathematical processing, retention or removal of some part of the shape information (e.g. tooth asymmetry). However, the magnitude of the effects, and the fact that they actually induce a dissimilarity in the results, is heavily dependent on the analysis performed (e.g. DA, PCA).

Thus, in statistical classifications, results are similar and major dissimilarity effects are connected with size (although in this case this dissimilarity is normal since it reveals a difference between form and shape descriptors) and with the elimination of some part of the shape information when Fourier amplitudes are used. Using complex Fourier coefficients instead of linear measurements does not greatly improve the gain

Table 6. Difference in intraspecific variance based on the reduced morphospace and associated probabilities. OD, difference in observed Total Variance between *Microtus agrestis* and *M. arvalis* (positive values correspond to a larger multivariate dispersion of *M. agrestis*). $TV_{agres.}$ and $TV_{arv.}$ indicate the bootstrap estimator (200 replicates) of intraspecific multivariate variance of the two species (mean of the bootstrap sample with this standard deviation in brackets). *z*-values are based on bootstrap estimators. P_{rand} = probability based on sampled randomization (simple difference and *z*-statistic with 5000 replicates); P_{norm} = probability based on normal approximation for the *z*-statistic

Descriptor	OD	P_{rand}	$TV_{agres.}$	$TV_{arv.}$	*z*-value	P_{rand}	P_{norm}
Traditional	4.25	0.0472	9.99 [1.25]	5.88 [0.97]	2.60	0.0086	0.0093
Log Ratio	–0.45	0.8486	11.24 [1.26]	11.88 [1.53]	–0.32	0.7612	0.7467
CDFA A nstd	5.87	<0.0001	11.46 [0.93]	5.71 [0.67]	5.01	<0.0001	<0.0001
CDFA A std	3.52	0.0408	9.33 [0.84]	5.85 [0.84]	2.9	0.003	0.0034
CDFA C nstd	11.60	0.0094	26.03 [3.06]	14.65 [2.56]	2.85	0.006	0.0044
CDFA C std	6.95	0.1710	22.29 [2.95]	15.49 [2.93]	1.63	0.1112	0.1020
EFA A nstd	7.98	<0.0001	13.08 [2.33]	5.39 [0.68]	3.17	<0.0001	0.0015
EFA A std	4.17	0.0210	10.03 [1.00]	5.96 [0.88]	3.05	0.0038	0.0023
EFA C nstd	13.41	0.0184	27.00 [4.25]	14.08 [2.57]	2.60	0.0106	0.0092
EFA C std	7.73	0.1916	23.22 [3.56]	15.76 [2.75]	1.66	0.106	0.0976

in classification, and may even involve a loss when amplitudes or other mathematical processing (elliptic) are used.

In voles, particularly in the case of *Microtus*, optimization of the linear descriptor on the anterior part of the tooth seems quite adequate for discriminating between species. In keeping with the developmental pathway of vole teeth (Jernvall *et al*., 2000; Salazar-Ciudad & Jernvall, 2002), major evolutionary innovations and thus major features for separating taxa occur in all likelihood in the anterior part of the tooth. Thus, optimization on this part weights the descriptor in favour of potentially informative characters. By contrast, outline-based descriptors provide a more accurate description of shape, although the informative component has the same weight as the uninformative component.

Results are similar in morphospace analysis among descriptors. However, the major effect of dissimilarity is not the same as in statistical classification. The type of information extracted (linear vs. outline) has a greater influence in this application. Similarities among morphospaces based on specimen distribution show that the information extracted is the major source of dissimilarity. Thus, interspecific differentiation is well recovered from morphospaces based on differently extracted information, while intraspecific differentiation is not quantified in the same way and is markedly different. Between outline descriptors, segregation is still recovered from amplitude- and coefficient-based descriptors.

Problems appear to occur with size standardization of elliptic descriptors. However, there are considerable similarities with both aspects of inter- and intraspecific differentiation. As for statistical classification, the developmental pathway of vole teeth can go some way toward explaining these results. Major dental features inducing interspecific differentiation are contained in all descriptors. However, outline and Fourier coefficients in particular allow us to better quantify the shape aspect (the uninformative part in statistical classification). Mammalian teeth are acknowledged to be highly evolvable with small developmental changes inducing large changes in the size and number of small cusps (Jernvall, 2000). In all likelihood, the developmental process of iterative addition of lateral cusps in voles accounts for this evolvability of new small cusps in the most anterior part of the tooth. Such small cusps occur regularly in the anterior loops of the tooth in *M. agrestis*. Linear measurements did not take into account such shape changes, contrary to outline techniques. Thus, this aspect of the evolvability of new cusps is more fully described by outline descriptors, although it corresponds to a lower level of morphological differentiation (e.g. intraspecific) and explains the divergence of morphospaces based on the different information extracted.

Despite dissimilarities among morphospaces, patterns of variation are very similar. One major effect is due to the utilization of form or shape descriptors: conserving size induces a difference in variance between the two species. This difference is maintained when standardized amplitudes (in elliptic or complex methods) are used. However, this is difficult to explain. The results are more or less conservative depending on the test used, but also depending on the type of information extracted.

The impact of descriptors on results varies and is related to the analysis performed. Thus, linear-based descriptors appear to be very effective at capturing major features of shape, describing interspecific differentiation and performing species discrimination.

However, they fail when a more modular description of shape is needed, for example when a study focuses on more precise aspects of shape such as tooth evolvability.

The complex method of Fourier analysis appears to have a number of advantages over classical elliptic analysis. The complex formulation of outline means information can be synthesized and the signal processed with the covariation of the two real signals. Our analyses suggest that complex methods yield somewhat better results in many cases, such as a gain in classification. However, they are limited to 2-dimensional signals whereas the elliptic analysis may be extended to 3-dimensional ones (Lestrel, 1997c).

In conclusion, our study emphasizes that the concept of a good descriptor is relative to the problem being examined. Despite their different intrinsic qualities, two descriptors can yield similar results if the requisite information is contained in both.

ACKNOWLEDGEMENTS

We are particularly grateful to the Senckenberg Museum (Frankfurt am Main, Germany) for the loan of material. We thank A. Bertin, P. Neige and two anonymous reviewers for their constructive comments on earlier drafts and C. Sutcliffe for assistance with the English. This work is a contribution to the projects 'Différenciation et espèces' and 'Macroévolution et dynamique de la biodiversité', UMR CNRS 5561-Biogéosciences. N. N. was supported by a doctoral grant from the Ministère de l'Education et de la Recherche. All MATLAB functions used in the paper are available at http://www.u-bourgogne.fr/BIOGEOSCIENCE/NNavarro/Navarro.htm.

REFERENCES

Adams DC, Rohlf FJ. 2000. Ecological character displacement in Plethodon: biomechanical differences found from a geometric morphometric study. *Proceedings of the National Academy of Sciences, USA* **97:** 4106–4111.

Airoldi J-P, Flury BD, Salvioni M. 1995. Discrimination between two species of *Microtus* using both classified and unclassified observations. *Journal of Theoretical Biology* **177:** 247–262.

Bertin A, David B, Cézilly F, Allibert P. 2002. Quantification of sexual dimorphism in *Asellus aquaticus* (Crustacea: Isopoda) using outline approaches. *Biological Journal of the Linnean Society* **77:** 523–534.

Brunet-Lecomte P. 1988. Les campagnols souterrains (Terricola, Arvicolidae, Rodentia) actuels et fossiles d'Europe occidentale. Unpublished Dphil Thesis, Université de Bourgogne.

Carrasco MA. 2000. Species discrimination and morphological relationships of kangaroo rats (*Dipodomys*) based on their dentition. *Journal of Mammalogy* **81:** 107–122.

Chaline J. 1972. *Les rongeurs du Pleistocène moyen et supérieur de France (Systématique – biostratigraphie – paléoclimatologie).* Paris: Edition du CNRS.

Chaline J. 1987. Arvicolid data (Arvicolidae, Rodentia) and evolutionary concepts. In: Hecht MK, Wallace B, Prance GT, eds. *Evolutionary biology.* New York: Plenum Press, 237–310.

Chaline J, Brunet-Lecomte P, Campy M. 1995. The last glacial/interglacial record of rodent remains from the Gigny karst sequence in the French Jura used for palaeoclimatic and palaeoecological reconstructions. *Palaeogeography, Palaeoclimatology and Palaeoecology* **117:** 229–252.

Chaline J, Brunet-Lecomte P, Montuire S, Viriot L, Courant F. 1999. Anatomy of the arvicoline radiation (Rodentia): Palaeogeographical, palaeoecological history and evolutionary data. *Annales Zoologici Fennici* **36:** 239–267.

Chaline J, Laurin B. 1986. Phyletic gradualism in a European Plio-Pleistocene *Mimomys* lineage (Arvicolidae, Rodentia). *Paleobiology* **12:** 203–216.

Chaline J, Laurin B, Brunet-Lecomte P, Viriot L. 1993. Morphological trends and rates of evolution in arvicolids (Arvicolidae, Rodentia): towards a punctuated equilibria/disequilibria model. *Quaternary International* **19:** 27–39.

Conroy CJ, Cook JA. 2000. Molecular systematics of a holarctic rodent (*Microtus*: Muridae). *Journal of Mammalogy* **81:** 344–359.

Conroy CJ, Demboski JR, Cook JA. 1999. Mammalian biogeography of the Alexander Archipelago of Alaska: a north temperate nested fauna. *Journal of Biogeography* **26:** 343–352.

Conroy CJ, Hortelano Y, Cervantes FA, Cook JA. 2001. The phylogenetic position of southern relictual species of *Microtus* (Muridae: Rodentia) in North America. *Mammalian Biology* **66:** 332–344.

Corti M, Fadda C, Simson S, Nevo E. 1996. Size and shape variation in the mandibule of fossorial rodent *Spalax ehrenbergi*. In: Marcus LF, Corti M, Loy A, Naylor GJP, Slice DE, eds. *Advances in morphometrics*. New York: Plenum, 303–320.

Courant F, Brunet-Lecomte P, Volobouev V, Chaline J, Quere J-P, Nadachowski A, Montuire S, Bao G, Viriot L, Rausch R, Erbajeva M, Shi D, Giraudoux P. 1999. Karyological and dental identification of *Microtus limnophilus* in a large focus of alveolar echinococcosis (Gansu, China). *Comptes Rendus de l'Academie des Sciences, Series III, Sciences de la Vie* **322:** 473–480.

Creighton GK. 1980. Static allometry of mammalian teeth and the correlation of tooth size and the body size in contemporary mammals. *Journal of Zoology, London* **191:** 435–443.

Della-Maggiore V, Sekuler AB, Grady CL, Bennet PJ, Sekuler R, McIntosh AR. 2000. Corticolimbic interactions associated with performance on a short-term memory task are modified by age. *Journal of Neuroscience* **20:** 8410–8416.

Dommergues C-H. 2001. *CDFT complex discrete fourier transform (Matlab package)*, Version 2.7. Dijon: Biogéosciences-Dijon, UMR CNRS 5561.

Dommergues E, Dommergues J-L, Magniez F, Neige P, Verrecchia EP. 2003. Geometric measurement analysis versus Fourier series analysis in shape characterization

using the gastropod shell (Trivia) as an example. *Mathematical Geology* **35:** 887–894.

Eble GJ. 2000a. Contrasting evolutionary flexibility in sister groups: disparity and diversity in Mesozoic atelostomate echinoids. *Paleobiology* **26:** 56–79.

Eble GJ. 2000b. Theoretical morphology: state of the art. *Paleobiology* **26:** 520–528.

Eble GJ. 2002. Developmental morphospaces and evolution. In: Crutchfield JP, Schuster P, eds. *Evolutionary dynamics*. Oxford: Oxford University Press, 35–65.

Efron B, Tibshirani RJ. 1993. *An introduction to the bootstrap*. New York: Chapman & Hall.

Ferson S, Rohlf FJ, Koehn RK. 1985. Measuring shape variation of two-dimensional outlines. *Systematic Zoology* **34:** 59–68.

Foote M. 1990. Nearest-Neighbor analysis of trilobite morphospace. *Systematic Zoology* **39:** 371–382.

Foote M. 1991a. Morphological and taxonomic diversity in a clades's history: the blastoid record and stochastic simulations. *Contributions from the Museum of Paleontology, University of Michigan* **28:** 101–140.

Foote M. 1991b. Morphological patterns of diversification: examples from trilobites. *Palaeontology* **34:** 461–485.

Foote M. 1992. Rarefaction analysis of morphological and taxonomic diversity. *Paleobiology* **18:** 1–16.

Foote M. 1993. Contributions of individual taxa to overall morphological disparity. *Paleobiology* **19:** 403–419.

Foote M. 1994. Morphological disparity in Ordovician-Devonian crinoids and the early saturation of morphological space. *Paleobiology* **20:** 320–344.

Foote M. 1995. Morphological diversification of Paleozoic crinoids. *Paleobiology* **21:** 273–299.

Foote M. 1999. Morphological diversity in the evolutionary radiation of Paleozoic and post-Paleozoic crinoids. *Paleobiology* **25:** 1–115.

Gårding L. 2000. A simple model for interplay of predators, rodents and foot. *Journal of Theoretical Biology* **206:** 73–80.

Gingerich PD, Smith BH, Rosenberg K. 1982. Allometric scaling on the dentition of primates and prediction of body weight from tooth size in fossils. *American Journal of Physical Anthropology* **58:** 81–100.

Giraudoux P, Delattre P, Habert M, Quéré JP, Deblay S, Defaut R, Duhamel R, Moissenet MF, Salvi D, Truchelet D. 1997. Population dynamics of fossorial water vole (*Arvicola terrestris scherman*): a land use and landscape perspective. *Agriculture, Ecosystems and Environment* **66:** 47–60.

Hurth E, Montuire S, Schmittbull M, Schaaf A, Viriot L, Chaline J. 2004. Examination of the morphospace of three *Mimomys* lineages (Arvicolinae, Rodentia) by elliptical Fourier methods. *Colloquios de Paleontologia* vol. ext. **1:** 325–334.

Jaarola M, Tegelström H, Fredga K. 1999. Colonization history in Fennoscandian rodents. *Biological Journal of the Linnean Society* **68:** 113–127.

Jackson DA. 1993. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology* **74:** 2204–2214.

Jernvall J. 2000. Linking development with generation of novelty in mammalian teeth. *Proceedings of the National Academy of Sciences, USA* **97:** 2641–2645.

Jernvall J, Keränen SVE, Thesleff I. 2000. Evolutionary modification of development in mammalian teeth: Quantifying gene expression patterns and topography. *Proceedings of the National Academy of Sciences, USA* **97:** 14444–14448.

Kiefte M. 1999. *Discriminant Analysis Toolbox*, Version 0.3. Available from http://www.mathworks.com/matlabcentral

Kincaid DT, Schneider RB. 1983. Quantification of leaf shape with a microcomputer and Fourier transform. *Canadian Journal of Botany* **61:** 2333–2342.

Klingenberg CP, Badyaev AV, Sowry SM, Beckwith NJ. 2001. Inferring developmental modularity from morphological integration: Analysis of individual variation and asymmetry in bumblebee wings. *American Naturalist* **157:** 11–23.

Klingenberg CP, Zaklan SD. 2000. Morphological integration between developmental compartments in the *Drosophila* wing. *Evolution* **54:** 1273–1285.

Kuhl FP, Giardina CR. 1982. Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing* **18:** 236–258.

Lance RF, Kennedy ML, Leberg PL. 2000. Classification bias in discriminant function analyses used to evaluate putatively different taxa. *Journal of Mammalogy* **81:** 245–249.

Laplana C, Montuire S, Brunet-Lecomte P, Chaline J. 2000. Révision des *Allophaiomys* (Arvicolinae, Rodentia, Mammalia) des Valerots (Côte-d'Or, France). *Geodiversitas* **22:** 255–267.

Lebart L, Morineau A, Piron M. 2000. *Statistique Exploratoire Multidimensionnelle*. Paris: Dunod.

Legendre S. 1989. Les communautés de mammifères du Paléogène (Eocène supérieur et Oligocène) d'Europe occidentale: structures, milieux et évolution. *Münchner Geowissenschaftliche Abhandlunden A* **16:** 1–110.

Lestrel PE. 1997a. *Fourier descriptors and their applications in biology*. Cambridge: Cambridge University Press.

Lestrel PE. 1997b. Introduction. In: Lestrel PE, ed. *Fourier descriptors and their applications in biology*. Cambridge: Cambridge University Press, 3–21.

Lestrel PE. 1997c. Introduction and overview of Fourier descriptors. In: Lestrel PE, ed. *Fourier descriptors and their applications in biology*. Cambridge: Cambridge University Press, 22–44.

Lindström J, Ranta E, Kokko H, Lundberg P, Kaitala V. 2001. From arctic lemmings to adaptive dynamics: Charles Elton's legacy in population ecology. *Biological Reviews* **76:** 129–158.

Lobaugh NJ, West R, McIntosh AR. 2001. Spatiotemporal analysis of experimental differences in event-related potential data with partial least squares. *Psychophysiology* **38:** 517–530.

Lohmann GP, Schweitzer PN. 1990. On eigenshape analysis. In: Rohlf FJ, Bookstein FL, eds. *Proceedings of the Michigan morphometrics workshop.* Ann Arbor: University of Michigan Museum of Zoology Special Publication no. 2, 145–166.

MacLeod N. 1999. Generalizing and extending the eigen-

shape method of shape space visualization and analysis. *Paleobiology* **25:** 107–138.

Marcus LF, Corti M, Loy A, Naylor GJP, Slice DE. 1996. *Advances in morphometrics*. New York: Plenum.

McGhee GR. 1999. *Theoretical morphology: the concept and its applications*. New York: Columbia University Press.

McIntosh AR, Bookstein FL, Haxby JV, Grady CL. 1996. Spatial pattern analysis of functional brain images using partial least squares. *Neuroimage* **3:** 143–157.

McIntosh AR, Gonzales-Lima F. 1998. Large-scale functional connectivity in associative learning: Interrelations of the rat auditory, visual, and limbic systems. *Journal of Neurophysiology* **80:** 3148–3162.

McLellan T, Endler JA. 1998. The relative success of some methods for measuring and describing the shape of complex objects. *Systematic Biology* **47:** 264–281.

Mehlman DW, Shepherd UL, Kelt DA. 1995. Bootstrapping principal component analysis – a comment. *Ecology* **76:** 640–643.

Moellering H, Rayner JN. 1981. The harmonic analysis of spatial shapes using Dual Axis Fourier Shape Analysis (DAFSA). *Geographical Analysis* **13:** 64–77.

Moellering H, Rayner JN. 1982. The Dual Axis Fourier Shape Analysis of closed cartographic forms. *Cartographic Journal* **19:** 53–59.

Montuire S, Michaux J, Legendre S, Aguilar JP. 1997. Rodents and climate. I: a model for estimating past temperatures using arvicolines (Mammalia: Rodentia). *Palaeogeography, Palaeoclimatology, Palaeoecology* **128:** 187–206.

Musser GG, Carleton MD. 1993. Familly Muridae. In: Wilson DE, Reeder DM, eds. *Mammal species of the world: a taxonomic and geographic reference*. Washington DC: Smithsonian Institute Press, 501–576.

Néraudeau D, Viriot L, Chaline J, Laurin B, van Kolfschoten T. 1995. Morphological trends and rates of evolution in arvicolids (Arvicolidae, Rodentia): towards a punctuated equilibria/disequilibria model. *Palaeontology* **38:** 77–85.

Nestor PG, O'Donnell BF, McCarley RW, Niznikiewicz M, Barnard J, Shen ZJ, Bookstein FL, Shenton ME. 2002. A new statistical method for testing hypotheses of neuropsychological/MRI relationships in schizophrenia: partial least squares analysis. *Schizophrenia Research* **53:** 57–66.

Ripley BD. 1996. *Pattern recognition and neural networks*. Cambridge: Cambridge University Press.

Rohlf FJ. 1986. Relationships among eigenshape analysis, Fourier analysis, and analysis of coordinates. *Mathematical Geology* **18:** 845–854.

Rohlf FJ. 1993. *NTSYS-Pc, numerical taxonomy and multivariate analysis system*. New York: Exeter Software.

Rohlf FJ, Archie JW. 1984. A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera: Culicidae). *Systematic Zoology* **33:** 302–317.

Rohlf FJ, Corti M. 2000. Use of two-block partial least-squares to study covariation in shape. *Systematic Biology* **49:** 740–753.

Salazar-Ciudad I, Jernvall J. 2002. A gene network model accounting for development and evolution of mammalian teeth. *Proceedings of the National Academy of Sciences, USA* **99:** 8116–8120.

Sampson PD, Streissguth P, Barr HM, Bookstein FL. 1989. Neurobehavioral effects of prenatal alcohol: Part II. Partial least squares analysis. *Neurotoxicology and Teratology* **11:** 477–491.

Schmittbull M, Allenbach B, Schaaf A, Le Minor JM. 1997. Méthodes de Fourier elliptiques appliquées à la quantification de la forme des contours complexes. *Biométrie Humaine et Anthropologie* **15:** 233–241.

Sokal RR, Rohlf FJ. 1995. *Biometry: the principles and practice of statistics in biological research*, 3rd edn. New York: W.H. Freeman.

Sundberg FA. 1996. Morphological diversification of Ptychopariida (Trilobita) from the Marjumiid biomere (Middle and Upper Cambrian). *Paleobiology* **22:** 49–65.

Tabachnick RE, Bookstein FL. 1990. The structure of individual variation in Miocene *Globorotalia*. *Evolution* **44:** 416–434.

Wills MA, Briggs DEG, Fortey RA. 1994. Disparity as an evolutionary index. A comparison of Cambrian and Recent arthropods. *Paleobiology* **20:** 93–110.