



## Research Article

# Log-shape ratios, Procrustes superimposition, elliptic Fourier analysis: three worked examples in R

Julien CLAUDE<sup>a,\*</sup><sup>a</sup>Institut des Sciences de l'Evolution de Montpellier, cc064, Université de Montpellier 2, 2 Place Eugène Bataillon, 34095 Montpellier cedex 5, France.

## Keywords:

rodents  
morphometrics  
*Rattus*  
Asia  
R

## Article history:

Received: 18 June 2012

Accepted: 9 April 2013

## Acknowledgements

I thank Yannick Chaval who organized body shape measurements in the CERoPath database, and who performed molecular barcoding. The landmark coordinates were obtained by Miss Duangkamol Phoophitpong (Kasetsart University), while pictures of teeth were taken by Sylvie Agret; body shape measurements were obtained from the members of the CERoPath project. Arden Hulme-Beaman, Andrea Cardini, Anna Loy, Paolo Colangelo seriously improved the first versions of this manuscript and made valuable comments. This study is sponsored by the ANR 07BDIV012 CERoPath, I1-CEPL-002 BiodivHealthSEA, and ANR-11-BSV7-008 BigTooth projects.

## Abstract

This publication uses and presents R routines that perform various morphometric analysis in the context of rodent systematics. The morphological variation of two commensal rat species, *Rattus exulans* and *Rattus tanezumi*, is analysed and the potential for discrimination between the two is assessed. Specimens were trapped in three localities of Northern and North-Eastern Thailand. Shape and size variation are analysed in regards to sex, species, and geographical effects with various morphometric methods: log-shape ratios on body measurements, elliptic Fourier analyses on teeth outlines, Procrustes superimposition on skull coordinates. Both species are significantly different; however, the discrimination seems to be better on skull Procrustes coordinates and on teeth size than on other morphometric data set. Where different allometries exist between species and where species differ in size and shape, it is shown that filtering allometry using the approach of Burnaby (1966) can improve the discrimination between species. Sex size and shape dimorphism is reduced by comparison to interindividual variation. Shape variation varies between sampled localities for *Rattus exulans*, this is not the case for *Rattus tanezumi*. This pattern is possibly related to the more commensal life of *R. exulans*.

## Introduction

Exploring morphological variation and relating this variation to explanatory factors is the essential purpose of morphometrics. In systematics, morphometrics primarily helps to quantify and describe differences between taxa or populations (e.g., Loy et al. 1993; Cardini et al. 2009; Viscosi and Cardini 2011; Chiari and Claude 2012). At present, there is a very large tool box for obtaining morphometric variables that can be routinely analysed via statistical analyses (see for instance the list given at <http://life.bio.sunysb.edu/morph/>). Morphometric methods have undergone an important progression starting from “traditional” or “multivariate” morphometrics to “modern” or “geometric morphometrics” (Rohlf and Marcus, 1993; Adams et al., 2004). While “traditional” morphometrics mostly relies on collection of raw linear measurements, “geometric” morphometrics analyses collections of anatomical landmarks expressed as point coordinates to quantify shape and size. Depending on the diverse available methods, size and several shape variables can be extracted from the morphometric recording on a set of specimens (Claude, 2008). Among others, these can be simple transformation of linear measurements (log-shape ratios), more sophisticated variables such as Procrustes coordinates obtained through superimposition methods for landmark data, or coefficients of shape functions fitted to curves or surfaces (e.g. elliptic Fourier analyses). All these techniques have been applied in mammal evolution and systematics (see Corti et al. 2000), and they are now routinely used by researchers.

As afore mentioned, a wide range of software is available to perform morphometric procedures and perform statistical analyses, but few offer the possibility to extract diverse parameters of shape variation for statistical shape analyses with a unique language and software. Moreover, few of these software can be run in diverse operational sys-

tems, which means that, a “Windows” or an “Apple” user may not have access to the same programs. However, many of the morphometric techniques have been recently imported into the R language and environment (Claude, 2008). The R project started in the 90’s and now provides a series of advantages for obtaining several kinds of morphometric data and for analysing these morphometric data with a wide array of statistical analyses (R Core Team, 2013). R has many advantages on other software: it is free, it can be run under various operating system (Windows, Linux, Apple), it is evolving with the help of a large community of users and developers. Moreover, the R language is close to the statistical jargon, and an enormous amount of literature about R is now available, often freely on the web. In addition to the R core, a very diverse number of packages has been developed in various disciplinary fields that have strong affinities with shape analysis (for instance, packages performing phylogenetic or comparative analyses, analyses of ecological communities). There is also a tremendous number of statistical tools, that are usually used in modern systematics, ecology and evolution, like fixed effects and mixed effects linear modeling, multivariate statistics, circular data, spatialised data, phylogenetics, genetics, comparative data, and community analyses. R also has graphical interfaces that permit the production a very large array of graphics easily customisable by the users. Finally, some packages and several functions have been developed for performing morphometrics, and other are being developed. The packages available on the CRAN include *shapes* and *geomorph* for Procrustes methods (Adams and Otárola-Castillo, 2013; Dryden, 2013), *Momocs* for outline analyses (Bonhomme et al., 2012), *LOST* for missing morphometric data simulation and estimation (Arbour and Brown, 2012). In addition, the package *Morpho* (Schlager, 2013) for 3D analyses is also available at <http://morpho-rpackage.sourceforge.net/>; and the functions developed in Claude (2008) can be downloaded or sourced at <http://www.isem.univ-montp2.fr/recherche/files/2012/01/Rfunctions1.txt>. It is

\* Corresponding author

Email address: [julien.claude@univ-montp2.fr](mailto:julien.claude@univ-montp2.fr) (Julien CLAUDE)

therefore valuable to use R and develop it further, with the possibility to ultimately adopt it as standard. R comes at first with a command line, but some more interactive graphical interfaces with buttons have been developed for users who are not familiar with R language (the R commander GUI that can be installed with the Rcmdr package (Fox, 2005), for instance). The use of a command line may be at first daunting to the new comers. However, it also gives an enormous advantage, as you can write scripts, save them and keep in memory the whole analyses in a text file. This can represent a certain gain of time to the users and is more efficient than having to play with buttons; in addition, because scripts are written in R, they can be posted to other users which review and improve them or to adapt it to other data sets. While describing an original study, the purpose of this paper is also to provide simple R commands for several kinds of morphometric analyses to be used with a number of data sets: landmark coordinates, shape measurements, or outline coordinates. The codes are supplied as supplementary material. The functions and codes that were used here principally come from Claude (2008) and are explained in that book. The supplementary material also contains some functions of the newly developed packages for obtaining graphics and performing tests. For the reader who may not be familiar with R, “R for beginners” by Paradis (2005) provides the essentials.

The scripts presented here are focused on taxonomic and simple evolutionary questions regarding rodent phenotypic characteristics. The applied part of the study aimed at depicting differences between species, relating shape variation to geographical factors and sex dimorphism, and to compare the congruency of results between different shape features using R routines. I also tried to determine which of the shape features and analyses could offer the best discrimination between the two taxa that were analysed. The main idea remains to show that one can rapidly obtain nice results and graphics with few line commands in R. I applied various morphometric methods for quantifying differences between two species of obligatory or facultatively commensal rodents, which both occurs in South-East Asia. I also tested whether sexual dimorphism or geographical differences could explain this variation. The two species of interest are *Rattus exulans* and *Rattus tanezumi*. Both species can be found indoors, sometime in the same house. They are considered as pests (Aplin et al., 2011) and potential reservoirs for several zoonotic diseases (Aplin et al., 2011; Lerdthusnee et al., 2008). Overall, the two species are very similar in shape but can be differentiated based on their size and ecology. *Rattus exulans* is a small sized species that is exclusively commensal in Thailand, found in house and farms, while *Rattus tanezumi* is larger, more opportunistic, living both in houses and diverse habitats (plantations, forests, agricultural lands) (Lekagul and McNeely, 1988; Corbet and Hill, 1992). There is, however, a considerable overlap in ecology and size variation, and small *R. tanezumi* found in human habitations can be easily confused with *Rattus exulans*.

In the literature, the two species are also differentiated by their mammae formula (Lekagul and McNeely, 1988; Corbet and Hill, 1992), but mammae formula are sometimes difficult to observe, and some variation occurs. As for many other cases, morphometrics (traditional or modern) can help identify species in the field. Traditional morphometrics is regularly used for species recognition. For instance, in Lekagul and McNeely (1988) and in Corbet and Hill (1992), tables of measurements are given and are intended to help in identification of taxa. In rodent taxonomy, however, differences can be subtle, with only minor skull or teeth differences, and the statistical analysis of complex shapes becomes sometimes the only resource for measuring how much species differ, and to potentially offer clues for taxonomic diagnosis. Furthermore, a morphometric analysis can be complementary to molecular methods (e.g., Guillot et al. 2012; Pages et al. 2013, as it allows to quantify phenotypic variation in populations. When species niches or geographical range overlap, morphometric analyses allow to evidence character displacements that could explain differences between populations (see Adams and Rohlf 2000; Loy and Capanna 1998). Here I explored the variability of the two rodent species using different morphometric methods.

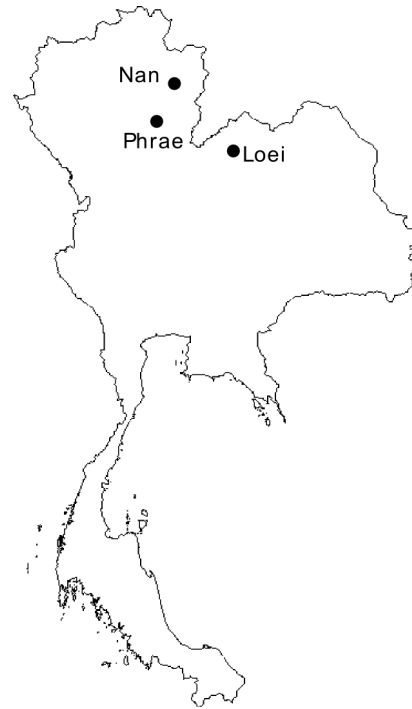


Figure 1 – Localities where rodents were sampled.

## Materials and Methods

Three original datasets have been obtained from 200 rodents trapped during field work in Northern and North-Eastern Thailand. Specimens were captured from three localities during field trips from 2006 to 2009 (Fig. 1, Tab. 1). The geographic range for a sampling site was around 20 km × 20 km, so that several habitats and houses were surveyed for each field session. Rodents were identified based on gross morphology, and some of them were identified via molecular analyses. *Rattus tanezumi* is often confounded in the literature with *Rattus rattus*, which was absent from the sites that were sampled for this study (Aplin et al., 2011; Lack et al., 2012; Pages et al., 2013). As there are taxonomic issues regarding *R. tanezumi* (Pages et al., 2013), most individuals of that species were also molecularly identified. *Rattus exulans* were determined mostly based on their gross morphology, mammae formula, and from the capture location (indoors). Whenever direct identification was not clear, or a morphological overlap was possible with *R. tanezumi* (for instance for subadults, unusually large specimens, when mammae formula was not clear, or when specimens were trapped outdoors) individuals were molecularly controlled also for this species. Here following are explanation of the three data sets (linear measurements, landmark coordinates on skulls and tooth outline coordinates) provided as supplementary material.

### Linear measurements

The first data-set “body.csv” is a compilation of five linear measurements (HBL: head + body length, EL: ear length, FL: hindfoot length, HL: head length, TL: tail length). Measurements were obtained using callipers (HL) and rulers (HBL, EL, FL, TL), following the protocol described in Auffray et al. (2011). The data set contains 10 variables organized in 10 columns. The first corresponds to the individual label, the three next are explanatory variables (sp: species, loc: locality (“n” for Nan, “l” for Loei, “p” for Phrae), sex: M for males, F for females);

Table 1 – Number of sampled individuals according to sexes, species and localities.

Locality	Loei		Nan		Phrae	
	F	M	F	M	F	M
<i>R. exulans</i>	22	19	20	19	23	19
<i>R. tanezumi</i>	12	22	10	9	11	13

the five next columns are response variables (morphometric measurements), and the last column indicates whether there was a molecular control on the species identification (either based on Cytochrome Oxidase 1 or on Cytochrome b). For simplicity and portability, this data set is written in .csv format and can be opened by any text editor. Data were collected only once by different operators. We did not test measurement error for these data, but are confident that our protocol is sufficiently standardized to limit variation due to the different operators.

Morphometric Procedures

The body measurements were analysed using the log-shape ratios approach (Mosimann, 1970). For each individual, size was computed as the geometric mean of all measurements, and each measurement was divided by size to obtain the shape ratios. The log of this quantity was used as raw data for subsequent analyses. Shape ratios are redundant in terms of information: one degree of freedom is lost due to scaling, and therefore four dimensions instead of five are necessary to describe shape variation. A principal components analysis was performed on five shape ratios and only the first four principal component scores were kept for multivariate analyses of variance. Observations with missing measurements (13.5%) were excluded from the analysis.

Statistical Analyses

In order to visualize how shape variation was structured, multivariate ordination of individuals was plotted on the first two PC axes using different colours and symbols for species, sex, and localities. Contributions of original variables on the corresponding eigenvectors (PC loadings) were analysed to understand which shape features had more influence on each PC.

The first PC axis represents 48.3% of shape variation while the second represents 18.9% (Fig. 2). The fourth PC (not shown here) seems to be related with species differentiation (*Rattus exulans* having higher scores), but there is an important overlap between species even on this axis. The fourth PC axis opposes the variables HBLand ELwith the variables FL, HLand FL, which means that *Rattus exulans* have in general relatively smaller ears and body length, and longer head, foot and tail compared to *Rattus tanezumii*.

Differences between Species, Sexes, and Localities

Effects of sexes, species and localities were estimated on geometric size using a multiple linear model, and the significance of these factors was tested using an ANOVA with type II sums of squares. Type II sums of

squares are calculated in such way that the effect of a factor, for a given order of interaction, is evaluated once all the other factors were taken into account. This approach does not violate the principle of marginality, although it is conservative (see Venables 1998; Claude et al. 2003, 2004). Similarly, a multiple and multivariate linear model was applied on principal components of shape variation with non-zero eigenvalues. The sex, species, locality factors, the size variable, as well as the interactions until the third order were considered as explanatory variables. A multivariate analysis of variance using type II sums of squares was performed on the different variances and covariances explained by the factors and covariables. Results indicate that size was significantly related to species but not to other factors (Tab. 2).

Table 2 – ANOVA on geometric size (loc: localities, sex: sex, sp: species).

Factor	SS	df	F	p-value
sex	3.6	1	0.21	0.65
loc	14.1	2	0.42	0.66
sp	9914.5	1	591.29	<0.0001
sex:loc	4.9	2	0.14	0.87
sex:sp	29.3	1	1.75	0.19
loc:sp	79.5	2	2.37	0.10
sex:loc:sp	8.1	2	0.24	0.78
Residuals	2682.8	160		

Size, species and localities significantly explained shape variation, while sex shape dimorphism was not significant (Tab. 3). No interaction of second or third order were found to be significant. Also, no interaction was found between species and size, suggesting that allometries are similar among groups.

By analysing each species separately (tests not shown in the main text), it can be seen that the locality effect is highly significant for *R. exulans* ( $p < 0.0001$ ) and only significant for *R. tanezumii* ( $0.01 < p < 0.05$ ), and that there is a triple interaction between sex, locality and size for *R. exulans*.

Table 3 – MANOVA on the first four PCs of log-shape ratios.

Factor	df	Pillai	approx. F	num. df	den. df	p-value
sex	1	0.04	1.46	4	147	0.22
loc	2	0.24	5.14	8	296	<0.0001
sp	1	0.47	33.21	4	147	<0.0001
size	1	0.56	46.9	4	147	<0.0001
sex:loc	2	0.05	0.90	8	296	0.51
sex:sp	1	0.03	1.29	4	147	0.28
sex:size	1	0.04	1.66	4	147	0.16
loc:sp	2	0.07	1.39	8	296	0.20
loc:size	2	0.09	1.76	8	296	0.08
sp:size	1	0.012	0.44	4	147	0.78
sex:loc:sp	2	0.07	1.34	8	296	0.22
sex:loc:size	2	0.07	1.28	8	296	0.25
sex:sp:size	1	0.01	0.32	4	147	0.86
loc:sp:size	2	0.06	1.15	8	296	0.33

Performance of Log-Shape Ratios for species Identification

In order to evaluate whether it is possible to easily distinguish species based on shape variables, a linear discriminant analysis using the species as group factor was performed on the log-shape ratios of molecularly identified specimens. Although the two species were significantly different in shape, it was not possible to obtain a good discrimination on log-shape ratios. The predictive discriminant analysis based on molecularly identified specimens correctly assigned only 25% of the specimens that were not molecularly controlled. This extremely low percentage (less than the worse expected: 50%) probably comes from the unbalanced sampling of the reference (few *R. exulans* were molecularly identified which may distort the discriminant coefficients in favour to special features of this small group).

A linear discriminant analysis using the species grouping was also performed on the whole log-shape ratio dataset. The percentage of correctly assigned specimens was computed using a leave-out-one cross validation procedure (jackknife) and only reached 73%.

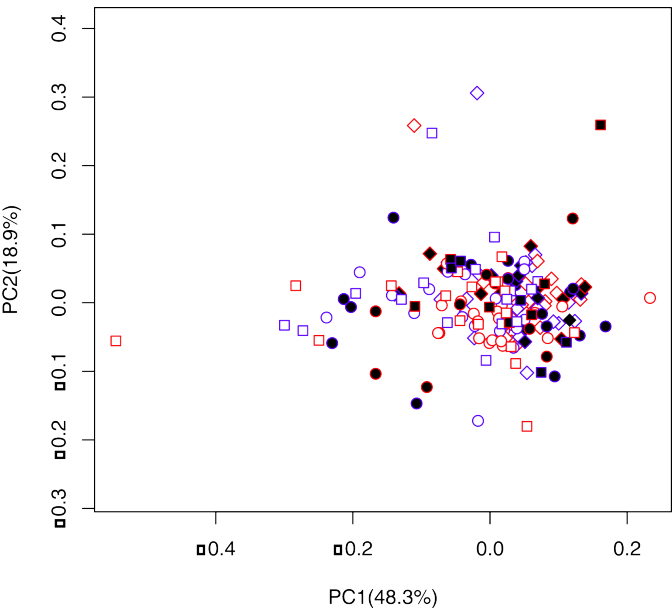
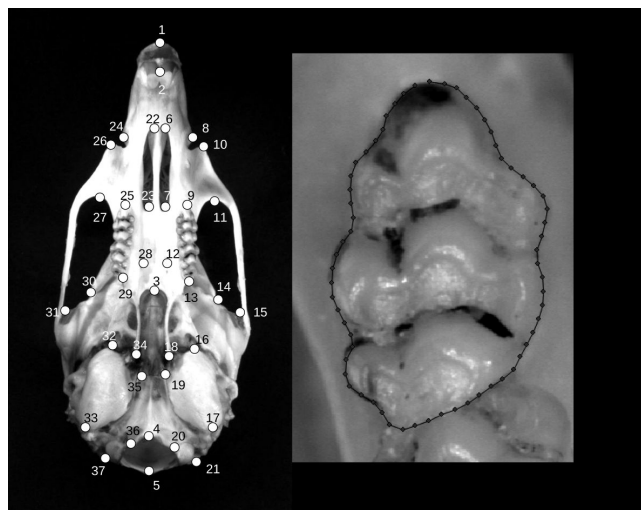


Figure 2 – PCA on log-shape ratios, localities are displayed with different symbols (circle, square, diamond), symbol outline colours represent sexes (red: females, blue: males), while symbol inner colours represent species (white: *R. exulans*, black: *R. tanezumii*).



**Figure 3** – Left: Landmark locations on the palatine view of the rodent skull (the two landmarks on the scale are not represented). Right: Upper first molar with digitized points around the outline.

Finally, when a discriminant analysis was directly applied on body shape measurements (therefore, when size was included), the percentage of correctly classified individuals reached 98% for the non-molecularly identified specimens. The same percentage of correct assigned individuals were obtained when using the leave-one-out cross-validation on the whole data set. By introducing this new variable, the linear combination of variables corresponding to discriminant coefficients was therefore much more efficient to discriminate between species.

### Landmarks and Procrustes superimposition

A second dataset, “skull.tps”, was produced from skulls. This data set is a collection of 39 points digitized for appraising the variation of the skull in palatine view (Fig. 3). The dataset contains the same individuals as the linear measurement dataset except one specimen. This dataset was obtained by first photographing the skulls of the specimens with a Pentax K200 camera, keeping always the same focal and distance between the specimens and the camera. Specimens were repositioned, rephotographed, and digitized a second time in order to estimate the percentage of error measurements during the digitization process. For each configuration, the 37 first pairs of coordinates correspond to the landmarks digitized on the skull palatine view (Fig. 3). Two additional points, spaced by 1 cm from each other, were measured on a millimetre scale photographed together with the skull, and used for scaling objects. For this study, landmark coordinates were originally recorded using the TPSdig2 digitization software (Rohlf, 2013), but could have been directly digitized using R functions of the package *geomorph*, such as `digitize2d` or a more general function like `locator`. The .tps format is explained in Rohlf (2013). Each image was labelled so that the four first characters corresponded to the specimen number, the fifth to the species, the sixth to the locality, the seventh to the sex, and the eight indicated whether the specimen was identified with molecular data (g) or directly on the field (f); the last character corresponds to the session number.

### Morphometric Procedures

Coordinates of landmarks were transformed from pixel into cm by dividing the raw coordinates by the Euclidean distance between the two landmarks digitized on the scale. All configurations (including replicates) were scaled to unit centroid size, translated, rotated, and optimally superimposed through the Generalized partial Procrustes Analysis (Dryden and Mardia, 1998; Claude, 2008). In this procedure, symmetric and asymmetric components were not partitioned. Details concerning Procrustes superimpositions can be found in Bookstein (1990, 1991, 1996); Goodall (1991); Small (1996); Rohlf and Slice (1990); Dryden and Mardia (1998); Viscosi and Cardini (2011). Centroid size

was used as an estimator for size. Procrustes coordinates were projected into the Euclidean tangent shape space using an orthogonal projection (Kendall, 1984; Goodall, 1991; Small, 1996; Dryden and Mardia, 1998; Claude, 2008). Superimposed coordinates are redundant: there are more coordinates than the number of dimensions in the shape space because the translation, scaling and rotation consume two, one and one degrees of freedom respectively (Dryden and Mardia, 1998). Therefore, a principal components analysis was performed on superimposed coordinates, and the 2 (dimensions) by 37 (landmarks) minus 4 (lost degrees of freedom) PCs were considered for multivariate analyses of variance.

### Measurement Error

Percentage of error measurement was obtained following the ANOVA approach described in Yezerinac et al. (1992), directly on centroid size and it was adapted to Procrustes data. For Procrustes coordinates, I followed the Procrustes ANOVA approach described in Goodall (1991); Klingenberg and McIntyre (1998); Claude et al. (2003). The among and within variances were calculated directly from the mean squares and crossproducts corresponding to the specimen and residual sources of variation. The percentage of measurement error is less than 1% for centroid size, and 26% for shape. The specimen factor is always significant, which means that interindividual size and shape variations are stronger than variation between replicated measurements on the same individual. One can note that this percentage of measurement error could be greater if one consider a smaller stratum for our samples (e.g., one single sex, one single species, one single locality).

### Principal Components of Shape Variation

A principal component analysis was performed on averaged configurations for each individual (average of the two replicates) and ordination of individuals was explored on the first two PCs. To understand which shape features were involved in the patterns of variation observed along the two axes, reconstruction of extreme morphologies along each PC was obtained for Procrustes data. The projection of individuals on eigenvectors (Fig. 4) shows that both species are well distinguished along PC 2, which involves the length of the molar row relatively shorter for *Rattus exulans* than for *Rattus tanezumi*, and the orientation of the incisor. Although PCs are computed in such way that each PC is independent and orthogonal to the others, one can see that there is a relationship between PC 1 and PC 2 if one consider each species independently (Fig. 4).

These apparent relationships certainly come from the fact that strong allometric relationships exist (cf. patterns observed on PC 1 that remind rodent skull ontogeny) and possibly differ between species (see statistical analyses below).

### Differences between Species, Sexes, and Localities

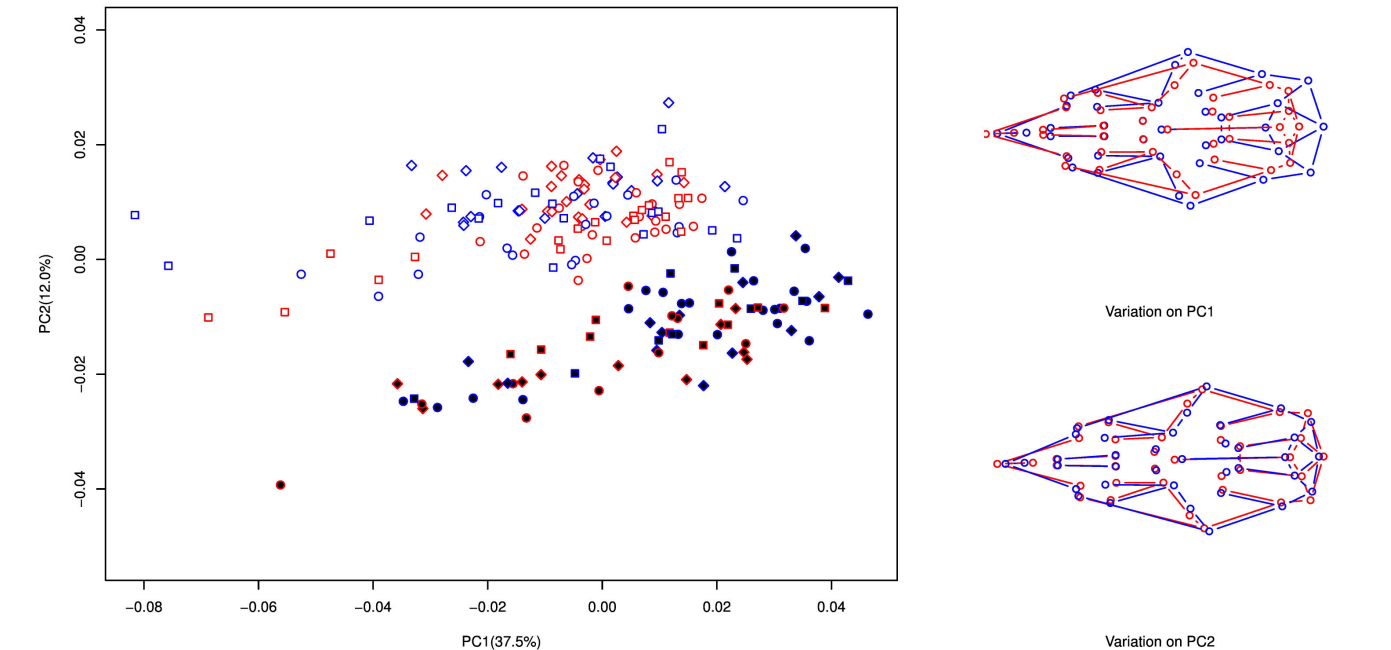
As for log-shape ratios, a univariate linear model on centroid size was applied to understand the possible effects of species, locality and sex. Similarly, a multivariate linear model was applied on principal components of shape variation with non-zero eigenvalues. The sex, species, locality factors, the size variable, as well as the interactions until the third order were considered in this model as explanatory variables. Univariate (for size) and multivariate (for shape) analyses of variance using type II sums of squares were performed on the different variances explained by the factors and covariables. Skull centroid size was influenced by species, and sex size dimorphism differed between species (Tab. 4). Males of *Rattus tanezumi* appear larger than females, while sex size dimorphism was not detected in *Rattus exulans*.

Skull shape variation was significantly affected by species, size, sexes and localities. In addition, size significantly interacted with species, and species with localities (Tab. 5). The patterns of shape differences between localities differed in the two species and the allometric growth pattern differed among species.

### Allometry-free approach

Because size interacted with species in the previous linear model, an allometry-free approach was conducted following the Burnaby Procedure (Burnaby, 1966). This approach permitted to filter out the effect of





**Figure 4** – PCA on Procrustes coordinates, similar symbols of Fig. 2. Patterns of variation along PC 1 and PC 2 are presented on the right side of the plot, blue corresponding to minimal scores, red to maximal ones.

**Table 4** – ANOVA on skull centroid size (loc: localities, sex: sex, sp: species).

Factor	SS	df	F	p-value
sp	146.086	1	652.94	<0.0001
sex	0.770	1	3.44	0.07
loc	0.599	2	1.34	0.26
sp:sex	1.803	1	8.06	0.0050
sp:loc	0.001	2	0.00	0.99
sex:loc	0.226	2	0.50	0.60
sp:sex:loc	0.588	2	1.31	0.27
Residuals	41.839	187		

**Table 5** – MANOVA on the 37 × 2 – 4 first shape principal components for Procrustes data.

Factor	df	Pillai	approx. F	num. df	den. df	p-value
sp	1	0.91	16.44	70	108	< 0.0001
size	1	0.92	17.77	70	108	< 0.0001
sex	1	0.51	1.63	70	108	0.0108
loc	2	1.21	2.38	140	218	< 0.0001
sp:size	1	0.58	2.10	70	108	0.0003
sp:sex	1	0.35	0.83	70	108	0.80
sp:loc	2	0.94	1.39	140	218	0.0147
size:sex	1	0.42	1.12	70	108	0.29
size:loc	2	0.87	1.20	140	218	0.11
sex:loc	2	0.73	0.89	140	218	0.78
sp:size:sex	1	0.40	1.03	70	108	0.44
sp:size:loc	2	0.83	1.11	140	218	0.24
sp:sex:loc	2	0.62	0.70	140	218	0.99
size:sex:loc	2	0.64	0.73	140	218	0.98

growth from the data. Although type II sums of squares are used for estimating effects, it could be interesting to remove all variation that could be due to growth. In order to reach this goal, I projected the data onto a space where ordination of individuals for every species would be independent of growth, rather than analyse variation once the effect of size was taken into account through regression (this is actually what is done with type II sums of squares). Since each species displayed different allometries, this approach is completely different than analysing variation after applying a linear model on individuals introducing size, and interaction between size and species. In practice, the vector of allometric coefficients for each species was obtained, and Procrustes coordinates were projected onto a an orthogonal space based on the direction of these vectors following the procedure of Burnaby (1966).

Once allometric growths within species were filtered from the shape variation, the two species appeared well discriminated on the first transformed PC axis (Fig. 5). In this allometry-free shape space, as for the previous analysis, species differentiation is related to the relative length of the tooth row and the orientation of the incisor. These characters can therefore be used independently of specimen size (ultimately age).

Species, sexes, and localities significantly differed in this new allometry-free shape space (Tab. 6). There is an interaction between the locality and species factors indicating that differences between localities are not the same for both species, once growth is filtered out from the data.

By examining each species separately (data not shown but see tests in supplementary material), *Rattus exulans* skull shape significantly differed between localities, with a marginal effect of sex, while *Rattus tanezumi* skull shape did not differ between localities and sexes.

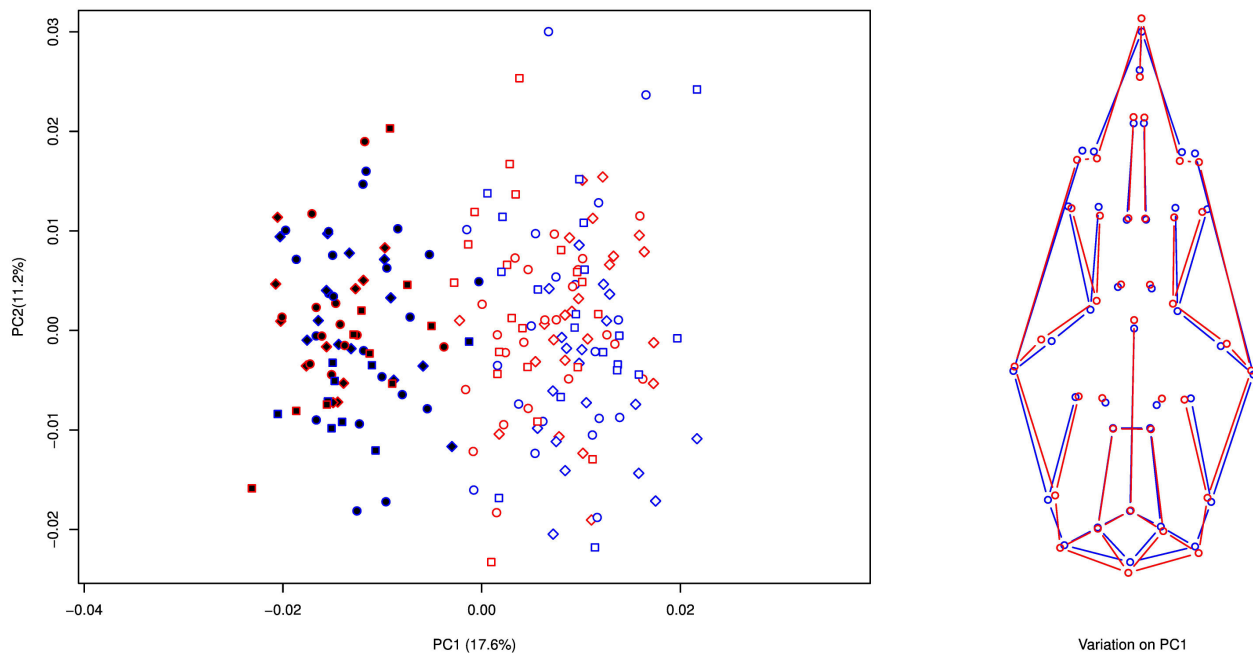
**Performance of Procrustes Data for Species Identification and Species Differences**

A linear discriminant analysis using the species factor was performed on the Procrustes coordinates for the molecularly identified specimens in order to know whether one can predict the species based on Procrustes data. The prediction of the specimens that were not molecularly controlled on this analysis was weak: only 26.8% of these were assigned to the correct species. This low score can in part be explained by the unbalanced samples in the reference or by the confounding effect of growth.

A linear discriminant analysis using the species was then performed on the dataset of Procrustes coordinates including all specimens, and the percentage of correctly assigned specimens was computed using a

**Table 6** – MANOVA on the 37 × 2 – 6 first allometry free shape principal components for Procrustes data.

Factor	df	Pillai	approx. F	num. df	den. df	p-value
sex	1	0.47	1.55	68	120	0.0190
sp	1	0.95	33.56	68	120	< 0.0001
loc	2	1.14	2.37	136	242	< 0.0001
sex:sp	1	0.40	1.20	68	120	0.20
sex:loc	2	0.70	0.95	136	242	0.62
sp:loc	2	0.91	1.47	136	242	0.0046
sex:sp:loc	2	0.70	0.95	136	242	0.62



**Figure 5** – PCA on Procrustes coordinates corrected for intraspecific ontogenetic allometries, similar symbols of Fig.2. Patterns of variation along PC 1 are presented on the right side of the plot, blue corresponding to minimal scores, red to maximal ones.

leave-one-out cross-validation (jackknife) procedure. In this analysis, 99.5% of correct re-assignment was obtained.

Discriminative features between species were estimated by rescaling the linear discriminant coefficients and are presented in Fig. 6. Indeed the metric of the discriminant space corresponds to the Mahalanobis distance and it is desirable to re-incorporate the pooled intra-group variance on the coefficients in order to depict shape changes in terms of original units (here displacements of landmarks in cm). As expected from the observations on PCAs, shape features that better discriminate between species concerned the relative length of the tooth row, the incisor position, as well as differences in the morphology of the posterior part of the skull (Fig. 6).

When intraspecific allometries are filtered out, 92.6% of individuals are correctly assigned to the good species for the samples that were not identified by molecular markers; while 99.5% of individuals are correctly re-assigned using the leave-one-out cross-validation procedure on the whole data set.

### Elliptic Fourier Analysis

“teeth.tps” is a collection of points digitized on the outline of the first right upper molar (Fig. 3) recorded a subset of 62 individuals. The teeth were photographed with a CCD camera mounted on a stereomicroscope. For each specimen, the two first digitized coordinates correspond to two points defined by an inframillimetric scale. As for skulls, specimens were randomly repositioned, photographed and digitized a second time for later estimating measurement error due to the

digitization process. Points were digitized along the outline of the teeth in a clockwise way starting from the anterior end of the first upper molar using TPSdig2 (Rohlf, 2013). Sixty four points were resampled from these former set using TPSdig2 (Rohlf, 2013). In R, such a procedure can be done using the `locator` function and/or more automatised functions such as those presented in Claude (2008). In future years, it can be expected that some improvements of the current packages will offer a GUI for easily digitizing curves. Images were labelled in a similar way as the “skull.tps” dataset.

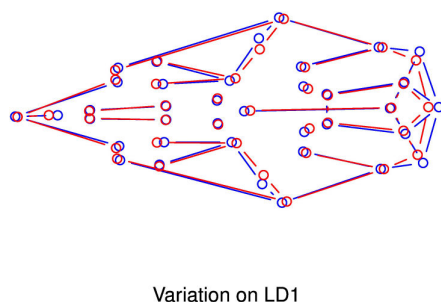
### Morphometric Procedures

As for skull coordinates, pixels were transformed to mm by dividing the raw coordinates by the Euclidean distance of the scale and multiplying by the scale actual length. For each configuration, the set of coordinates was then described using normalized elliptic Fourier coefficients. Details of the method can be found in Kuhl and Giardina (1982); Rohlf and Archie (1984); Claude (2008). Size was estimated as the longer radius of the first fitting ellipse. To reduce the number of variables by comparison to the number of individuals, only the first harmonics that showed a reasonable digitization error rate (< 35%) were kept. Indeed high order harmonics are more sensitive to small random variations (and therefore to noise) that could be introduced during the digitization process.

### Measurement Error

Sixty-four points were sampled for elliptic Fourier analysis, but only the first 32 harmonics were retained (following the Nyquist theorem Shannon 1949). The percentage of error on harmonic coefficients is calculated with a similar approach as the Procrustes ANOVA. The mean sums of squares were first calculated for the four coefficients of each harmonics to observe the evolution of percentage of error according to the rank of harmonics. The total measurement error rate was obtained by summing the different mean squares for the four coefficients within each harmonics. Tooth size measurement error (based on the larger radius of the first ellipse) reaches 1.4%. Tooth shape measurement error depends on harmonic rank (Fig. 7).

The first normalised harmonic coefficients displayed an important digitization error rate (61%) while the coefficients of the six following reach around 30%. The high level of error found for the first harmonic certainly comes from the variation in orientation of the tooth under the stereomicroscope, since the corresponding coefficients measure the width on length ratios (and this ratio depends on the orientation



**Figure 6** – Shape features that best discriminate species: “typical” *Rattus tanezumi* shape in red, while “typical” *Rattus exulans* is in blue.

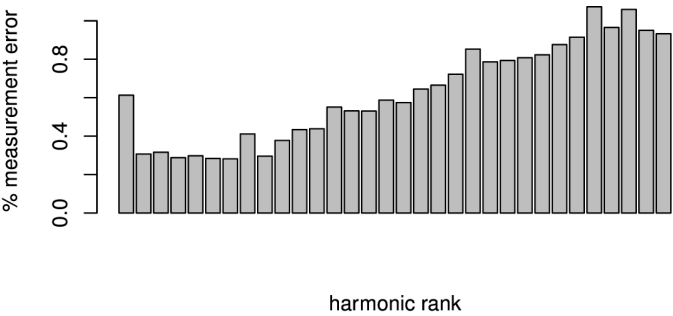


Figure 7 – Change of measurement error in regard to harmonic rank.

of the occlusal surface relative to the horizontal plan). After the seventh harmonic, the percentage of error variation increased. Shape variables that were considered for further tests were therefore summarized by the seven first harmonic coefficients by excluding the first harmonic coefficients.

Analysis of Principal Components of Shape Variation

As for Procrustes data, a principal components analysis was performed on averaged elliptic Fourier coefficients of tooth outlines for each individual. Ordination of individuals was later plotted on the first two PC axes. To understand which shape features were involved, reconstruction of extreme morphologies along each PC was obtained for elliptic Fourier data. In order to reconstruct tooth outline from Fourier coefficients, the inverse Fourier transform was used to reconstruct these theoretical teeth outlines (Rohlf and Archie, 1984; Claude, 2008). A large overlap between species is observed for tooth outline shape variation on the first two principal component axes (Fig. 8).

Differences between Species, Sexes, and Localities

The effects of sex, species and locality were estimated by a linear model on tooth size. Tooth size was expressed by the length of the major axis of the best fitting ellipse defined by the first harmonic coefficients. A multivariate linear model was applied on principal components of the normalized elliptic Fourier coefficients with non-zero eigenvalues. The sex, species, locality factors, the variable size, as well as the interactions until the third order were considered as explanatory variables. Univariate and multivariate analyses of variance using type II sums of squares were performed on the different variances explained by the factors and covariables for size and shape respectively. Species significantly differed in size and there was no overlap in tooth size between species (Tab. 7). No other factors contributed significantly as a source of size variation for teeth. Species also significantly differed in shape and there was a significant relationship between tooth shape and size (Tab. 8). However, no interaction and no sex effects were significant.

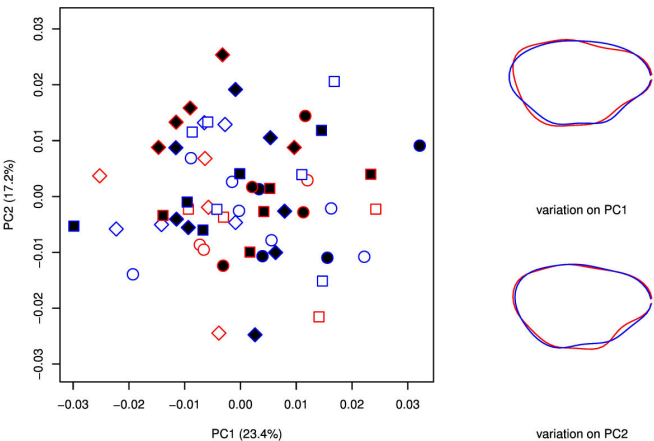


Figure 8 – PCA on the Fourier coefficients obtained from tooth outline, patterns of variation along PC 1 and PC 2 are presented on the right side of the plot, blue corresponding to minimal scores, red to maximal ones.

Table 7 – ANOVA on on first upper molar size.

Factor	SS	df	F	p-value
sex	0.00002	1	0.1480	0.70
sp	0.13466	1	1196.65	< 0.0001
loc	0.00046	2	2.04	0.14
sex:sp	0.00013	1	1.17	0.28
sex:loc	0.00009	2	0.41	0.67
sp:loc	0.00048	2	2.13	0.13
sex:sp:loc	0.00005	2	0.20	0.82
Residuals	0.00540	48		

Table 8 – MANOVA on the elliptic Fourier coefficients.

Factor	df	Pillai	approx. F	num. df	den. df	p-value
sp	1	0.88	4.47	24	15	0.0021
sex	1	0.78	2.23	24	15	0.06
loc	2	1.38	1.47	48	32	0.12
size	1	0.87	4.36	24	15	0.0024
sp:sex	1	0.45	0.50	24	15	0.94
sp:loc	2	1.10	0.82	48	32	0.74
sp:size	1	0.72	1.57	24	15	0.18
sex:loc	2	1.23	1.07	48	32	0.43
sex:size	1	0.41	0.44	24	15	0.97
loc:size	2	1.18	0.96	48	32	0.56
sp:sex:loc	2	1.17	0.94	48	32	0.58
sp:sex:size	1	0.44	0.49	24	15	0.94
sp:loc:size	2	1.30	1.24	48	32	0.27
sex:loc:size	2	1.19	0.98	48	32	0.54

When species were tested separately on the 17 first PCs (99.5% of shape variation) for size, sex and locality and the different interaction between these factors, significant differences were found between localities for *R. exulans* but not for *R. tanezumini*. Sex had a marginal effect only for *R. tanezumini*.

Performance of Elliptic Fourier Coefficients for Species Identification and Species Differences

As too few individuals of *Rattus exulans* were molecularly identified, only the predictive discriminant analysis based on the whole dataset of coefficients was performed with the leave-one-out cross-validation procedure. Sixty-six percent of individuals were correctly reclassified in this analysis. Linear discriminant coefficients were rescaled by re-incorporating intragroup variance covariance to identify shape features that increased the “ratio” of inter- on intra-specific variation. The shape features that discriminate between species were very subtle and are illustrated in Fig. 9. It can be seen that the anterior part of the tooth is better demarcated in *R. tanezumini* (Fig. 9).

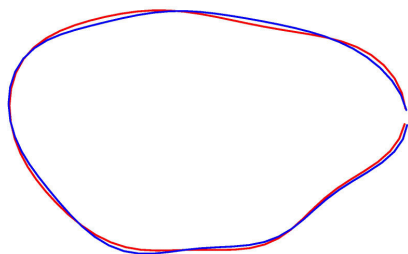
Discussion

Both species differed for size, linear measurements, skull and tooth shapes. However, some structures more effectively discriminated between species: tooth size as well as features of skull palatine views seem better discriminators than log-shape ratios or tooth shape parameters. Discrimination between species mostly involved tooth row length.

*R. exulans* is smaller than *R. tanezumini* in teeth, skull, and body measurements. The length of the tooth row relative to the skull size is smaller in *R. exulans* than in *R. tanezumini*, suggesting evolutionary allometries between species. Post-natal ontogeny strongly structure the morphological variation of the skull shape within each species, but the relationships between shape and size differ between species. Not only mean shape changes occurred in the evolution of *Rattus* but developmental features, such as the allometric relationships between shape and size were also modified. It is interesting to note that tooth outline shape is rather preserved in the two species, while tooth size differs. Tooth shape has often been found to differ between closely related species within the Murinae (e.g., Macholan 2006).

Discriminating between closely related rodent species can sometimes be difficult. The *Rattus* genus is known to be a very diverse group for which delimiting species is difficult due to an important intraspe-





**Figure 9** – Tooth shape variation that best discriminates between species, the red outline corresponds to *R. tanezumi* phenotype while the blue outline corresponds to *R. exulans*.

cific variation (Rowe et al., 2011). When one can have access to the dynamic of growth and when this effect can be filtered out from variation, landmark data on skulls performs rather well to discriminate species. As there is also a good differentiation between species in size, adding this variable in the discriminant analysis can increase the discrimination between species. Based on body measurements, form (shape + size) discriminates rather well between the two species compared to shape only. This study also shows that equal sampling between groups as well as sufficient effective size matter for obtaining better predictions (see the very low score of predictive discriminant analyses when species have unequal numbers of observations).

All the shape features that could differentiate the two species were not exhaustively explored in this study. An exploration of other skull views, osteological traits, body measurements, or teeth could possibly yield also good results. In addition, it is clear now that sampling size should be large enough for morphometrics to serve at identifying specimen using predictive discriminant analyses.

In contrast to the analysis of shape ratios, Procrustes or outline analyses provide the considerable advantage to visually display shape changes along exploratory axes (PCA) or explanatory axes (linear discriminant axes). Candidate features for discriminating the two species include tooth length, configuration of the posterior part of the skulls (relatively wider in *R. exulans*).

Sexual dimorphism was significant only for skull shape and seems to be stronger in *R. exulans* than in *R. tanezumi*. No sexual dimorphism of size could be identified. One must, however, stress that significance of effects depends not only on the magnitude of effect variation but also on the residual variation. This characteristic if tested with controlled genetic and environmental variation could become possibly significant.

*Rattus exulans* differs between localities for skull morphometrics and tooth shape, while *Rattus tanezumi* does not. Motokawa et al. (2004) and Pages et al. (2013) reported morphological differences for *Rattus exulans* and for *Rattus tanezumi* at a larger geographical scale. The observed differences probably demonstrate that geographical variation exists at finer geographical scale for *Rattus exulans*. Although it is speculative at this stage, one could hypothesize that differences in demographic features and kinship structures may possibly explain the differentiation between *R. exulans* populations and the homogeneity between *R. tanezumi* populations. It is also possible that due to its strong relationship with humans, some individuals of *R. exulans* have settled a distinct population helped with human activity (transportation, importation of food, etc.). It is also possible that *R. tanezumi* is more panmictic because it exploits a larger range of habitats (indoor, agricultural and forested areas) than *R. exulans*, which is always commensal in Thailand. Larger sampling and molecular studies could probably help to understand the different results obtained for both species.

Finally, scripts that were developed for this study can easily be reused and adapted for other groups, and for addressing similar questions: morphological differentiation, evaluation of sex dimorphism, geographical variation. Furthermore, since morphometrics is now going along with several studies such as quantitative genetics, morphological integration, phylogenetics, fluctuating asymmetries, etc., and because many R libraries are developed and devoted with some aspects of these problems (mixed linear modelling, matrix calculation, reconstruction of ancestral character states), the script can probably serve as a base or as a tutorial for going beyond taxonomic issues. ☺

## References

- Adams D.C., Otárola-Castillo E., 2013. *geomorph*: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution* 4: 393–399. [R package version 1.0] <http://cran.r-project.org/web/packages/geomorph/index.html>
- Adams D., Rohlf F.J., 2000. Ecological character displacement in *Plethodon*: biomechanical differences found from a geometric morphometric study. *Proceedings of the National Academy of Sciences, U.S.A.* 97: 4106–4111.
- Adams D.C., Rohlf F.J., Slice D.E., 2004. Geometric morphometrics: ten years of progress following the “revolution”. *Int. J. Zool.* 71: 5–16.
- Aplin K.P., Suzuki H., Alejandro A., Chinen A.A., Chesser R.T., ten Have J., Donnellan S.C., Austin J., Frost A., Gonzales J.P., Herbreteau V., Catzeffis F., Soubrier J., Fang Y.-P., Robins J., Matisoo-Smith E., Bastos A.D.S., Maryanto I., Sinaga M.H., Denys C., Van Den Bussche R., Conroy C., Rowe K., Cooper A., 2011. Multiple geographic origins of commensalism and complex dispersal history of black rats. *Plos ONE*, 6(11): e26357. doi:10.1371/journal.pone.0026357
- Arbour J., Brown C., 2012. *LOST*: Missing morphometric data simulation and estimation. [R package version 1.0] <http://cran.r-project.org/web/packages/LOST/index.html>
- Auffray J.-C., Blasdel K., Bordes F., Chabé M., Chaisiri K., Charbonnel N., Chaval Y., Claude J., Cosson, J.-F., Dei-Cas E., Desquesnes M., Dobigny G., Douangboupba B., Galan M., Haukialmi V., Henttonen H., Herbreteau V., Hugot J.-P., Jiyoung T., Latanne A., Michaux J., Milocco C., Morand S., Pagès M., Phoophitpong D., Pumhom P., Ribas Salvador A., Soonchan S., Suputtamongkol Y., Waengsothorn S., Waywa D., Xuéreb A., 2011. *Protocols for Field and Laboratory Rodent Studies*. Kasetsart University Press.
- Bonhomme V., Picq S., Claude J., 2013. *Momocs*: Shape Analysis of Outlines. [R package version 0.2-02] <http://vincentbonhomme.fr/momocs/>
- Bookstein F.L., 1990. Introduction to methods for landmark data. In: Rohlf F.J., Bookstein F.L. (Eds.) *Proceedings of the Michigan Morphometric Workshop*. University of Michigan Museum of Zoology Special Publication 2. 215–226.
- Bookstein F.L., 1991. *Morphometric tools for landmark data: Geometry and Biology*. Cambridge University Press, Cambridge.
- Bookstein F.L., 1996. Combining the tools of geometric morphometrics. In: Marcus L.F., Corti M., Loy A., Naylor G.J.P., Slice D., Bookstein F.L. (Eds.) *Advances in Morphometrics*. NATO ASI Series. Plenum Press, New York. 131–152.
- Brown C., Arbour J., Jackson D., 2012. Testing of the effect of missing data estimation and distribution in morphometric multivariate data analyses. *Systematic Biology* 61: 941–956.
- Burnaby T.P., 1966. Growth-invariant discriminant functions and generalized distances. *Biometrics* 22: 96–110.
- Cardini A., Nagorsen D., O’Higgins P., Polly P.D., Thorington J., Tongiorgi P., 2009. Detecting biological distinctiveness using geometric morphometrics: an example case from the Vancouver island marmot. *Ethology, Ecology and Evolution* 21: 209–223.
- Chiari Y., Claude J., 2012. Morphometric identification of individuals when there are more shape variables than reference specimens: a case study in galapagos tortoises. *Comptes rendus Biologies* 335: 62–68.
- Claude J., 2008. *Modern Morphometrics with R*. Springer, New York.
- Claude J., Paradis E., Tong H., Auffray J.-C., 2003. A geometric morphometric assessment of the effects of environment and cladogenesis on the evolution of the turtle shell. *Biological Journal of the Linnean Society* 79: 485–501.
- Claude J., Pritchard P., Tong H., Paradis E., Auffray J.-C., 2004. Ecological correlates and evolutionary divergence in the skull of turtles: a geometric morphometric assessment. *Systematic Biology* 53: 933–948.
- Corbet G.B., Hill J.E., 1992. *The Mammals of the Indomalayan Region: a systematic review*. Natural History Museum Publications, Oxford University Press, Oxford.
- Corti M., Marcus L., Hingst-Zaher E., 2000. Introduction to the Symposium: Geometric morphometrics in mammalogy. *Hystrix* 11(1): 3–7. doi:10.4404/hystrix-11.1-4133
- Dryden I.E., 2013. *shapes*: Statistical shape analysis. [R package version 1.1-6] <http://www.maths.nottingham.ac.uk/~ild/shapes>
- Dryden I.E., Mardia K.V., 1998. *Statistical Shape Analysis*. Wiley, Chichester.
- Fox J., 2005. *The R commander: A basic-statistics graphical user interface to R*. *Journal of Statistical Software*, 14(9): 1–42.
- Goodall C.R., 1991. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society, series B* 53: 285–339.
- Guillot G., Renaud S., Ledevin R., Michaux J., Claude J., 2012. A Unifying Model for the Analysis of Phenotypic, Genetic and Geographic Data. *Systematic Biology* 61(5): 897–911.
- Kendall D., 1984. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society* 16: 81–121.
- Klingenberg C.P., McIntyre G.S., 1998. Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with Procrustes methods. *Evolution* 54(4): 1363–1375.
- Kuhl F.P., Giardina C.R., 1982. Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing* 18: 236–258.
- Lack J.B., Greene D.U., Conroy C.J., Hamilton M.J., Braun J.K., Mares M.A., Van Den Bussche R.A., 2012. Invasion facilitates hybridization with introgression in the *Rattus rattus* species complex. *Molecular Ecology* 21(14): 3545–3561.
- Lekagul B., McNeely J.A., 1988. *Mammals of Thailand*, 2nd edition. Darnsutha Press, Thailand.
- Lerdthusnee K., Nigro J., Monkanna T., Leepitakrat W., Leepitakrat S., Insuan S., Charoensongsermkit W., Khilaimanee N., Akkagraisee W., Chayapum K., Jones J., 2008. Surveys of rodent-borne disease in Thailand with a focus on scrub typhus assessment. *Integrative Zoology* 3: 267–273.
- Loy A., Capanna E., 1998. A parapatric contact area between two species of moles (genus *Talpa*): character displacement investigated through the geometric morphometric of skull. *Acta Zoologica Academiae Scientiarum Hungaricae* 44(1–2): 151–164.
- Loy A., Corti M., Marcus L.F., 1993. Landmark data: Size and shape analysis in systematics. A case study on old world talpidae (Mammalia, Insectivora). In: Marcus L.F., Bello E., Garcia-Valdecasas A. (Eds.) *Contribution to Morphometrics*. Museo Nacional de Ciencias Naturales, Madrid. 215–240.



- Macholan M., 2006. A geometric morphometric analysis of the shape of the first upper molar in mice of the genus *Mus* (Muridae, Rodentia). *Journal of Zoology* 270: 672–681.
- Mosimann J.E., 1970. Size allometry: size and shape variables with characterizations of the lognormal and generalized gamma distributions. *Journal of the American Statistical Association* 65: 930–948.
- Motokawa M., Lin L.-K., Lu K.-H., 2004. Geographic variation in cranial features of the polynesian rat *Rattus exulans* (Peole, 1848) (Mammalia: Rodentia: Muridae). *The Raffles Bulletin of Zoology* 52(2): 653–663.
- Pages M., Bazin E., Galan M., Chaval Y., Claude J., Herbreteau V., Michaux J., Piry S., Morand S., Cosson J.-F., 2013. Cytonuclear discordance among southeast asian black rats (*Rattus rattus* complex). *Molecular Ecology* 22: 1019–1034.
- Pages M., Chaval Y., Herbreteau V., Waengsothorn S., Cosson J.-F., Hugot J.-P., Morand S., Michaux J., 2010. Revisiting the taxonomy of the Rattini tribe: a phylogeny-based delimitation of species boundaries. *BMC Evolutionary Biology* 10: 184.
- Paradis E., (2005). R for beginners. Available online at [http://cran.r-project.org/doc/contrib/Paradis-rdebuts\\_en.pdf](http://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf)
- R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>
- Rohlf F.J., 2013. tpsDIG2: Digitize landmarks & outlines from image files, scanner, or video. Available online at <http://life.bio.sunysb.edu/morph/soft-dataacq.html>
- Rohlf F.J., Archie A.W., 1984. A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera: Culicidae). *Systematic Zoology* 33(3): 302–317.
- Rohlf F.J., Marcus L.F., 1993. A revolution in morphometrics. *Trends in Ecology and Evolution* 8(4): 129–132.
- Rohlf F.J., Slice D.E., 1990. Extension of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology* 39: 40–59.
- Rowe K.C., Aplin K.P., Braverstock P.R., Moritz C., 2011. Recent and rapid speciation with limited morphological disparity in the genus *Rattus*. *Systematic Biology* 60: 188–203.
- Schlager S., 2013. Morpho: Calculations and visualizations related to Geometric Morphometrics. [R package version 0.17] <http://sourceforge.net/projects/morpho-rpackage/>
- Shannon C.E., 1949. Communication in the presence of noise. *Proc. Institute of Radio Engineers* 37(1): 10–21.
- Small C.G., 1996. The statistical theory of shape. Springer, New York.
- Venables W., 1998. Exegeses on linear models. Paper presented to the S-PLUS User's Conference Washington, DC, 8–9<sup>th</sup> October, 1998.
- Viscosi V., Cardini A., 2011. Leaf morphology, taxonomy and geometric morphometrics: A simplified protocol for beginners. *PLoS ONE* 6(10): e25630. doi:10.1371/journal.pone.0025630
- Yezerinac S.M., Loogheed S.C., Handford P., 1992. Measurement error and morphometric studies: statistical power and observer experience. *Systematic Biology* 41(4): 471–482.

Associate Editor: A. Loy