MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2011) **11**, 353–357 doi: 10.1111/j.1755-0998.2010.02924.x

COMPUTER PROGRAM NOTE

MORPHOJ: an integrated software package for geometric morphometrics

CHRISTIAN PETER KLINGENBERG
Faculty of Life Sciences, University of Manchester, Michael Smith Building, Manchester M13 9PT, UK

Abstract

Increasingly, data on shape are analysed in combination with molecular genetic or ecological information, so that tools for geometric morphometric analysis are required. Morphometric studies most often use the arrangements of morphological landmarks as the data source and extract shape information from them by Procrustes superimposition. The MORPHOJ software combines this approach with a wide range of methods for shape analysis in different biological contexts. The program offers an integrated and user-friendly environment for standard multivariate analyses such as principal components, discriminant analysis and multivariate regression as well as specialized applications including phylogenetics, quantitative genetics and analyses of modularity in shape data. MORPHOJ is written in Java and versions for the Windows, Macintosh and Unix/Linux platforms are freely available from http://www.flywings.org.uk/MorphoJ_page.htm.

Keywords: geometric morphometrics, multivariate statistics, phylogeny, Procrustes superimposition, quantitative genetics, shape

Received 19 July 2010; revision received 8 September 2010; accepted 10 September 2010

Morphometric analyses are in increasingly widespread use in conjunction with analyses of molecular data. Examples of such combined uses of data include mapping of shape data onto phylogenies inferred from DNA sequences, quantitative genetic analyses based on genealogies extracted from molecular marker information, or assessments of the effects of hybridization on shape asymmetry. Therefore, flexible tools for morphometric analysis are required by a growing community of users in various areas of ecology and evolutionary biology.

The most widespread approach in geometric morphometrics is to represent each specimen by the relative positions of morphological landmarks that can be located precisely and establish a one-to-one correspondence among all specimens included in the analysis. Shape is then defined as all the geometric information about a configuration of landmarks except for its size, position and orientation (Dryden & Mardia 1998). The shape information is extracted by a procedure called Procrustes superimposition, which removes variation in size, position and orientation from the data on landmark coordinates, and which is at the core of geometric morphometrics (Goodall 1991; Bookstein 1996; Dryden & Mardia 1998; Zelditch *et al.* 2004). The coordinates of the superimposed landmarks can be used in multivariate statistical analyses to address a wide range of biological questions (e.g. Klingenberg 2010).

The MORPHOJ software is based on this approach and aims to provide a flexible and user-friendly platform for a broad range of morphometric analyses for two- or three-dimensional landmark data. The goal is to provide a single, integrated environment for geometric morphometrics so that users can concentrate on the biological and statistical aspects of the analyses. As far as possible, the program automatically keeps track of properties such as the symmetry and dimensionality of landmark configurations and adjusts the analyses as needed. MORPHOJ implements the standard range of multivariate techniques that are widely used in geometric morphometrics as well as a number of more specialized or newer methods. This study introduces the MORPHOJ software and gives an overview of the methods it contains.

Implemented methods

After landmark data have been imported (from text files or files in the formats of other morphometric programs), the first step in a morphometric analysis is to extract shape information from the data with a Procrustes superimposition (Dryden & Mardia 1998). MORPHOJ implements a full Procrustes fit and projection onto the tangent

Correspondence: Christian Peter Klingenberg,
E-mail: cpk@manchester.ac.uk

space to the shape space (Dryden & Mardia 1998), which produces a new set of shape variables that can be used in further analyses. Information on the size of the landmark configuration is retained in the data set and available for subsequent analyses (centroid size and log-transformed centroid size; Dryden & Mardia 1998). After this initial step, a wide range of analyses can be used to analyse shape variation or to relate it to other information.

After extracting shape information, it is usually helpful to a search for outliers in the data, for which MorphoJ provides a graphical user interface (including the capability to repair mistakes where landmarks have been recorded in the wrong order). If the user has digitized specimens repeatedly, it is possible to quantify measurement error relative to the effects of biological interest by using Procrustes ANOVA (Klingenberg & McIntyre 1998; Klingenberg *et al.* 2002). This is particularly important if a study is focusing on subtle effects, such as variation within populations or left—right asymmetry.

The standard methods of multivariate statistics that are in widespread use in morphometrics are implemented in MorphoJ. Principal component analysis can be used to examine the main features of shape variation in a sample and as an ordination analysis for examining the arrangement of specimens in morphospace. Canonical variate analysis provides a different type of ordination analysis, which maximizes the separation of specified groups (species, ecotypes, etc.), and discriminant analysis with cross-validation indicates whether groups can be distinguished reliably. The patterns of shape variation can be compared between groups by matrix correlation and assessed statistically with matrix permutation tests, which have been adapted specifically to geometric morphometric data (Klingenberg & McIntyre 1998; Klingenberg *et al.* 2002).

Covariation of shape with other types of variables is an important aspect of morphometric studies, and MorphoJ offers several techniques that can be used in this context. Multivariate regression analysis can be used for assessing allometry or other relationships between variables, such as shape changes over time, which can be studied by regressing shape on size or on time (Monteiro 1999; Drake & Klingenberg 2008). Partial least-squares analysis is a method for studying associations between sets of variables and can be used in contexts such as ecomorphology (Adams & Rohlf 2000) or morphological integration (Klingenberg & Zaklan 2000; Klingenberg *et al.* 2001). It also can provide interesting insights concerning the covariation of shape and genetic markers (e.g. allele dosage of microsatellite markers; unpublished results).

Symmetric structures, such as vertebrate skulls, are frequently used in morphometric studies and pose some special challenges (Mardia *et al.* 2000; Klingenberg *et al.* 2002). MorphoJ separates the shape variation of such structures into components of symmetric and asymmetric variation, which provide information that is relevant to different biological questions (Klingenberg *et al.* 2002). The symmetric component represents the shape variation among individuals and is the focus for answering many biological questions, whereas the asymmetry component tends to be used in more specialized analyses, for instance as a measure of developmental instability in contexts such as hybridization (Mikula & Macholán 2008). For such uses, MorphoJ provides measures of individual asymmetry (Klingenberg & Monteiro 2005), which can be correlated with external variables such as ecological factors or measures of hybridization from molecular markers. Another use for data on morphological asymmetry, employed in a rapidly growing number of studies, is for characterizing the developmental basis of morphological integration (Klingenberg 2008).

MorphoJ offers several methods for studying morphological integration and modularity (Klingenberg 2008). In addition to general methods such as principal components, partial least squares and matrix correlation, which can be used for characterizing and comparing patterns of integration, MorphoJ also offers specialized methods for assessing hypotheses of modularity in landmark data (Klingenberg 2009).

Quantitative genetic studies are increasingly feasible, even in natural populations when pedigree information is available from behavioural observations or molecular markers (McGuigan 2006; Wilson *et al.* 2010), and the methods for such studies have been extended to shape analyses (Klingenberg & Leamy 2001; Klingenberg & Monteiro 2005; Myers *et al.* 2006; Klingenberg *et al.* 2010a). MorphoJ does not contain methods for estimating genetic covariance matrices and similar statistics, for which there are specialized software packages such as VCE (Groeneveld *et al.* 2008) or Wombat (Meyer 2007) that are freely available and can handle high-dimensional data as they are produced in geometric morphometrics, but MorphoJ can import genetic and environmental covariance matrices estimated by those packages and use them in specifically morphometric analyses. The current release of MorphoJ includes methods for finding shape variables of maximal or minimal heritability, useful for identifying genetic constraints on evolution, and for simulating hypothetical selection scenarios that can help in assessing genetic integration in a structure (Klingenberg & Leamy 2001; Martínez-Abadías *et al.* 2009; Klingenberg *et al.* 2010a). Methods for estimating selection on shape (Lande & Arnold 1983; Gómez *et al.* 2006) will be added in a new release in the near future.

MorphoJ contains methods for mapping shape data onto phylogenies using squared-change parsimony (Maddison 1991) and for comparative analyses such as independent contrasts (Felsenstein 1985). There is also a permutation test to establish whether a morphometric data set contains a phylogenetic signal (Klingenberg & Gidaszewski 2010). Phylogenies are imported into MorphoJ as NEXUS files (Maddison *et al.* 1997), as they are produced by most phylogenetic software and available from online databases such as Treebase (http://www.treebase.org).

User interface and data formats

The aim of MorphoJ is to provide a user-friendly and flexible environment for morphometric analyses. Accordingly, MorphoJ has a graphical user interface (Fig. 1) from which all analyses can be invoked. Multiple analyses of one or more data sets can be combined with each other to explore different aspects of the data. The data sets and analyses are organized into a project, which is intended to be a self-contained unit (e.g. the morphometric data and analyses that go into a paper

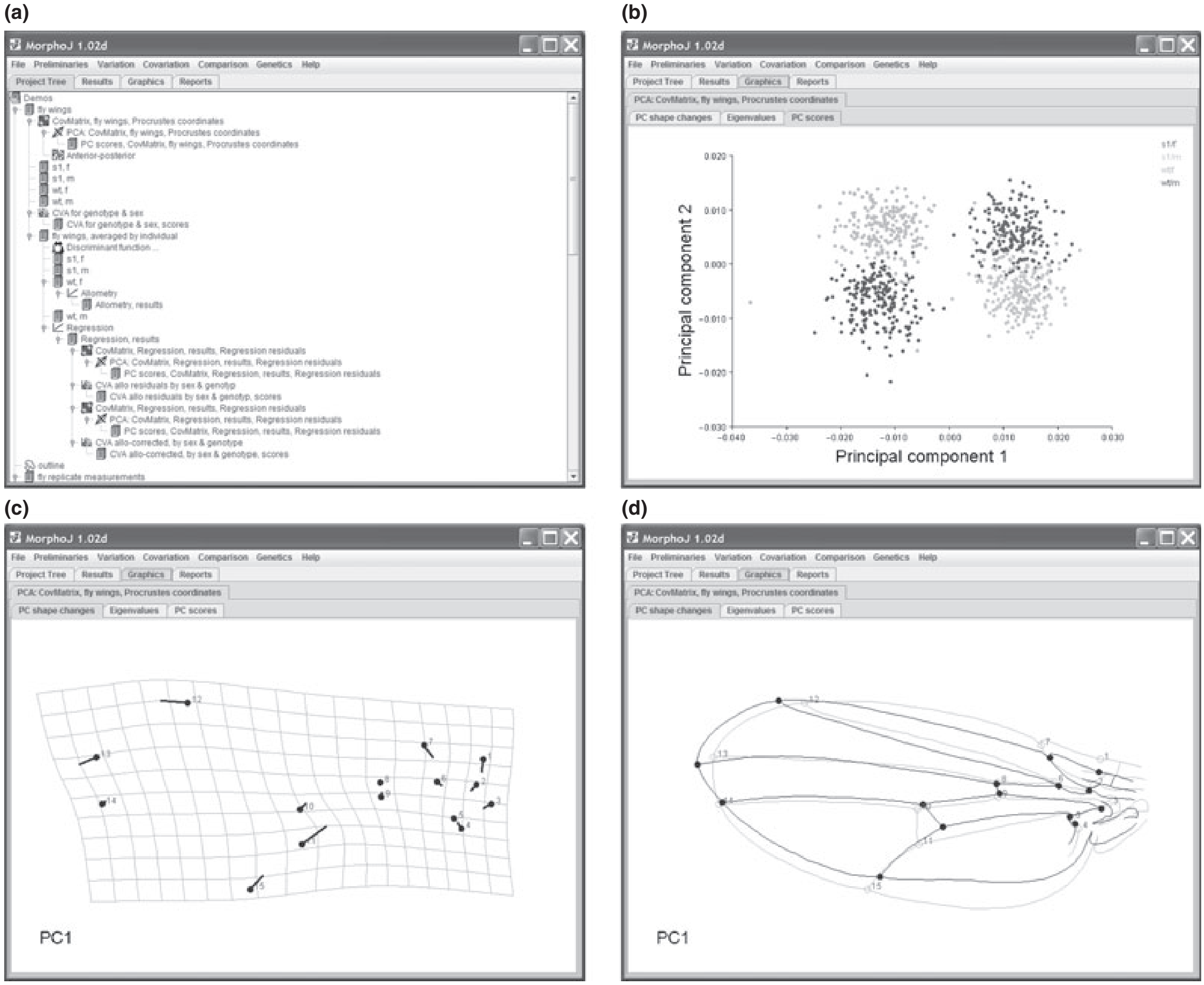

Fig. 1 User interface of the MorphoJ software (shown here for the Windows XP platform). The user interface consists of a series of menus containing various commands and four tabbed display areas with (a) The Project Tree tab showing the project and its contents as a tree structure. All data sets and analyses in the active project are shown in this tab, and results or graphical outputs can be invoked via pop-up menus. (b) A scatter plot of principal component scores, shown in the Graphics tab. MorphoJ provides various kinds of plots, including scatter plots and histograms, that present results from statistical analyses. (c) A transformation grid for visualizing a shape change (for the first principal component, in this case). (d) The same shape change as in (c), represented by a warped outline drawing (change from light to dark outline drawing). In addition to the possibilities shown in this figure, MorphoJ has several additional options for visualizing the results of morphometric analyses.

or thesis). Data sets inside a project can be related to each other by appropriate analyses (e.g. regression or partial least-squares analyses). The contents of the active project are displayed in the Project Tree window (Fig. 1a), where graphs and numerical results for each item can be called up by using pop-up menus. Projects can be saved to disc with all analyses and results they contain, so that work on a project can be taken up at a later time or the complete information can be exchanged among collaborators (projects are stored as XML files and are fully compatible across computer platforms).

MorphoJ can import raw landmark data from tab- or comma-delimited text files or from several of the customary file formats used in other morphometric software: the TPS series (http://life.bio.sunysb.edu/morph/soft-tps.html), NTSYSpc (Rohlf 2008) and Morphologika (O'Higgins & Jones 1998). MorphoJ exports data as tab-delimited text files, which can easily be imported into various spreadsheet or statistics programs. Additional information in the form of categorical data can be imported as ''classifiers'' and continuous variables can be included as ''covariates'', and are available for appropriate morphometric analyses. Also, it is possible to import results from analyses in other program packages (e.g. SAS, Matlab or R) as sets of shape changes for further analyses and visualization in MorphoJ.

MorphoJ provides various kinds of graphical outputs, including scatter plots and other standard types of graphs for visualizing statistical results (Fig. 1b). In addition, it provides several types of graphs for visualizing shape changes associated with the statistical results (Fig. 1c,d). For two-dimensional data, these include transformation grids (Fig. 1c) or warped outline drawings of the structure under study (Fig. 1d). For three-dimensional data, MorphoJ only provides basic visualizations of shape changes, but the information on shape changes can be exported to files that can be used by the Landmark software (Wiley *et al.* 2005) for presentation as morphed 3D surfaces (e.g. Drake & Klingenberg 2010; Klingenberg *et al.* 2010b). The graphs produced by MorphoJ can be exported in several graphics file formats, including vector graphics formats (including PDF, PostScript and SVG, which can be edited in software such as Inkscape or Adobe Illustrator to produce publication quality graphs) and several raster graphics formats (BMP, GIF, PNG).

Documentation for MorphoJ is provided as a User's Guide in HTML format, which is distributed with every download (the command ''User's Guide'' in the Help menu launches the computer's default browser to display the documentation) or available online (http://www.flywings.org.uk/MorphoJ_guide/).

Comparison to other morphometric software

Morphometrics software is currently available as standalone packages with graphical user interfaces or as packages of routines for programming environments such as Matlab or R (Claude 2008). This spectrum is associated with an inherent trade-off of user-friendliness versus flexibility. While programs with graphical user interfaces provide quick and convenient analyses, they cannot offer the flexibility of systems where users are required to do some of the programming themselves.

MorphoJ aims to take an intermediate position in this spectrum. It is based on a graphical user interface and therefore does not require users to have programming skills. But MorphoJ also aims to provide a maximum of flexibility by offering a wide range of analyses (often with several options, e.g. pooled within-group analyses) and an integrated user environment that facilitates combining different methods in the analysis of shape data.

Also, MorphoJ contains a number of unique features. It is currently the only program package that fully takes into account the symmetry of landmark configurations throughout the analyses—an important point because many biological structures, such as vertebrate skulls, are bilaterally symmetric. MorphoJ also contains some advanced tools for analyzing modularity and integration of shape.

Availability

MorphoJ is available freely under the Apache Licence, Version 2.0, from the web site http://www.flywings.org.uk/MorphoJ_page.htm. It is distributed as automatic installer software for Windows, Apple Macintosh and Unix/Linux platforms. Updates with bug fixes, improved documentation or additional morphometric methods are released occasionally.

Acknowledgements

I thank the members of my lab, various colleagues and the participants of my online course in morphometrics, who all helped me developing MorphoJ by making suggestions, giving advice and ''stress-testing'' the program with all sorts of data sets. Research in my lab, during the development of MorphoJ, was funded by BBSRC, the European Commission, the Leverhulme Trust, NERC, the Royal Society, and the Wellcome Trust.

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 of the USA*, **97**, 4106–4111.

Bookstein FL (1996) Biometrics, biomathematics and the morphometric synthesis. *Bulletin of Mathematical Biology*, **58**, 313–365.

Claude J (2008) *Morphometrics with R*. Springer, New York.

Drake AG, Klingenberg CP (2008) The pace of morphological change: historical transformation of skull shape in St. Bernard dogs. *Proceedings of the Royal Society of London, B Biological Sciences*, **275**, 71–76.

Drake AG, Klingenberg CP (2010) Large-scale diversification of skull shape in domestic dogs: disparity and modularity. *American Naturalist*, **175**, 289–301.

Dryden IL, Mardia KV (1998) *Statistical shape analysis*. Wiley, Chichester.

Felsenstein J (1985) Phylogenies and the comparative method. *American Naturalist*, **125**, 1–15.

Gómez JM, Perfectti F, Camacho JPM (2006) Natural selection on *Erysimum mediohispanicum* flower shape: insights into the evolution of zygomorphy. *American Naturalist*, **168**, 531–545.

Goodall CR (1991) Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society B*, **53**, 285–339.

Groeneveld E, Kovač M, Mielenz N (2008) *VCE user's guide and reference manual, version 6.0*. Institute of Farm Animal Genetics, Neustadt, Germany.

Klingenberg CP (2008) Morphological integration and developmental modularity. *Annual Review of Ecology, Evolution and Systematics*, **39**, 115–132.

Klingenberg CP (2009) Morphometric integration and modularity in configurations of landmarks: tools for evaluating a-priori hypotheses. *Evolution & Development*, **11**, 405–421.

Klingenberg CP (2010) Evolution and development of shape: integrating quantitative approaches. *Nature Reviews Genetics*, **11**, 623–635.

Klingenberg CP, Gidaszewski NA (2010) Testing and quantifying phylogenetic signals and homoplasy in morphometric data. *Systematic Biology*, **59**, 245–261.

Klingenberg CP, Leamy LJ (2001) Quantitative genetics of geometric shape in the mouse mandible. *Evolution*, **55**, 2342–2352.

Klingenberg CP, McIntyre GS (1998) Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with Procrustes methods. *Evolution*, **52**, 1363–1375.

Klingenberg CP, Monteiro LR (2005) Distances and directions in multidimensional shape spaces: implications for morphometric applications. *Systematic Biology*, **54**, 678–688.

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

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, Barluenga M, Meyer A (2002) Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. *Evolution*, **56**, 1909–1920.

Klingenberg CP, Debat V, Roff DA (2010a) Quantitative genetics of shape in cricket wings: developmental integration in a functional structure. *Evolution*, **64**, doi:10.1111/j.1558-5646.2010.01030.x.

Klingenberg CP, Wetherill LF, Rogers JL *et al.* (2010b) Prenatal alcohol exposure alters the patterns of facial asymmetry. *Alcohol*, doi: 10.1016/j.alcohol.2009.10.016.

Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution*, **37**, 1210–1226.

Maddison WP (1991) Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. *Systematic Zoology*, **40**, 304–314.

Maddison DR, Swofford DL, Maddison WP (1997) NEXUS: an extensible file format for systematic information. *Systematic Biology*, **46**, 590–621.

Mardia KV, Bookstein FL, Moreton IJ (2000) Statistical assessment of bilateral symmetry of shapes. *Biometrika*, **87**, 285–300.

Martínez-Abadías N, Paschetta C, de Azevedo S, Esparza M, González-José R (2009) Developmental and genetic constraints on neurocranial globularity: insights from analyses of deformed skulls and quantitative genetics. *Evolutionary Biology*, **36**, 37–56.

McGuigan K (2006) Studying phenotypic evolution using multivariate quantitative genetics. *Molecular Ecology*, **15**, 883–896.

Meyer K (2007) WOMBAT—A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *Journal of Zhejiang University Science B*, **8**, 815–821.

Mikula O, Macholán M (2008) There is no heterotic effect upon developmental stability in the ventral side of the skull within the house mouse hybrid zone. *Journal of Evolutionary Biology*, **21**, 1055–1067.

Monteiro LR (1999) Multivariate regression models and geometric morphometrics: the search for causal factors in the analysis of shape. *Systematic Biology*, **48**, 192–199.

Myers EM, Janzen FJ, Adams DC, Tucker JK (2006) Quantitative genetics of plastron shape in slider turtles (*Trachemys scripta*). *Evolution*, **60**, 563–572.

O'Higgins P, Jones N (1998) Facial growth in *Cercocebus torquatus*: an application of three-dimensional geometric morphometric techniques to the study of morphological variation. *Journal of Anatomy*, **193**, 251–272.

Rohlf FJ (2008) *NTSYSpc: Numerical Taxonomy System, ver. 2.20*. Exeter Publishing, Setauket, NY.

Wiley DF, Amenta N, Alcantara DA *et al.* (2005) Evolutionary morphing. *Proceedings of the IEEE Visualization 2005 (VIS'05)*, 431–438.

Wilson AJ, Réale D, Clements MN *et al.* (2010) An ecologist's guide to the animal model. *Journal of Animal Ecology*, **79**, 13–26.

Zelditch ML, Swiderski DL, Sheets HD, Fink WL (2004) *Geometric Morphometrics for Biologists: A Primer*. Elsevier, San Diego.