MODICOS: Morphometric and Distance Computation Software oriented for evolutionary studies.

A. Carvajal-Rodríguez (phD) and M. Gandarela Rodríguez

Departamento de Bioquímica, Genética e Inmunología, Facultad de Ciencias

Universidad de Vigo, 36310 VIGO, Spain (acraaj@uvigo.es)

ABSTRACT

Keywords: Size, shape, quantitative genetic variation.

Motivation: In the study of evolution is essential the consideration of quantitative differences as well as qualitative ones. Monitoring quantitative differences within and between populations requires the measurement of specific quantitative traits. Doubtless, size and shape are among such quantitative traits. The use of geometric morphometric techniques allow to distinguish shape and size in a precise and formal way. Hence, geometric morphometric distinguish between global and local changes in shape and also provides for allometry tests. At present there are some few freely available programs that perform morphometric computations, like tpsRelw. However, there is no program which allows for hierarchical structured data (e.g. families and individuals), or performs ANOVA analysis onto morphometric data in order to estimate genetic components of quantitative variation within and between populations.

Methods: MODICOS (MOrphometry and DIstance COmputation Software oriented for evolutionary studies) is a computer program with a user-friendly Windows 95/98/2000/NT/XP interface. MODICOS computes some morphometric measures as centroid size, uniform and non-uniform components of shape and use them to calculate estimates of quantitative (heritable)

variation both at intra and inter-population level. Specifically, MODICOS computes additive genetic variances of size and shape measurements both for half sib or full sib family design (one way ANOVA) and inter-population differentiation measure Q_{ST} (two way ANOVA).

Results: To show the use of MODICOS, a complete analysis (morphometric and genetic) was performed using a sample of two different ecotypes of *Littorina saxatilis* (rough periwinkle), each with 29 and 30 families respectively, and 3 individuals per family. It seems that the non-uniform shape component is more efficient capturing the effect of natural selection onto the two different ecotypes.

Availability:

http://webs.uvigo.es/acraaj/Modicos.htm

INTRODUCTION

The approach of geometric morphometric (Rohlf and Slice, 1990; Bookstein, 1991; Rohlf and Marcus, 1993; Rohlf and Bookstein, 2003) has proved its usefulness in the measure of biological form and the distinction between shape and size of morphological structures (Monteiro, Bordin and Furtado dos Reis, 2000 and references therein). Consequently, geometric morphometric has become increasingly important in genetics and evolutionary biology as suggested by an increasing number of publishing papers using these methods (Arnqvist and Thornhill, 1998; Cavalcanti *et al*., 1999; Douglas *et al*., 2001; Klingenberg *et al*., 2001; Palsson and Gibson, 2004; Carvajal-Rodríguez *et al*., 2005a).

The study of quantitative genetic variation both, within and between populations requires the measurement of specific quantitative traits. Doubtless, size and shape are among such traits. The advantage of geometric morphometric methods is that they allow distinguishing shape and size in a precise and formal way. Hence, geometric morphometric discriminate between global and local changes in shape and also provides for allometry tests. Thus, we will use geometric morphometric techniques for description of orthogonal components of variation, as these capture the major components of variation and are independent of observer bias (Bookstein, 1991).

It is worth mentioning that at present there are some few freely available programs that perform morphometric computations, like tpsRelw (Rohlf, 1998). However, as far as we know, there is no program which allows for hierarchical structured data (e.g. families and individuals), or performs ANOVA analysis onto morphometric data in order to estimate genetic components of quantitative variation within and between populations. Measuring the genetic variation contributing to variation in size and shape is crucial if one is interested in predicting trait evolutionary change (Falconer and MacKay, 1996). Furthermore, quantitative genetic variation of size and shape could

also be a useful tool to detect changes in biodiversity caused by different human induced activities (Carvajal-Rodríguez *et al*., 2005b). MODICOS (MOrphometry and DIstance COmputation Software oriented for evolutionary studies) performs two kinds of tasks, morphometric and genetics analysis, in a user-friendly and efficient way. For morphometric analysis, the input data consists in a series of *n* individuals, each represented by a set of *p* bi-dimensional landmark points, i.e. coordinate pairs *x*, *y* (see below). Size and shape measures are calculated from these data using geometric morphometric techniques. The output is written in a text file in a convenient format for posterior analysis with standard statistical packages. Furthermore, the output is also maintained in memory so that the morphometric measures linked to each individual could be immediately used as phenotypes to estimate genetic components of variances at the intra-population (heritabilities) and inter-population level (Q_{ST}, a dimensionless measure of quantitative genetic variance among populations) under an adequate experimental design. Such design assumes a hierarchical structure of data distributed in populations, families within populations and individuals within families. MODICOS also allows for the genetic analysis of user-provided traits. In this later case each individual in the input file consists in a single phenotypical value rather than a set of points.

METHODOLOGY

Data Input

Two different file formats are allowed. The input file for morphometric analysis should have the header [MODICOS - POINTS] while the input for genetics should have [MODICOS - PHENOTYPES]. Under any of these headers the first individual is identified as three numbers separated with points, e.g. 111.1.0 identifies population 111, family 1 and individual 0. Under each individual identifier, a column of *p* landmarks (pairs *x, y* separated by comma or space) should

follow in the case of a morphometric input file. In the case of genetic input file a single phenotype follows under each individual identifier. After the *p* landmarks or the single phenotype, the next individual identifier should appear with its corresponding landmarks or single phenotype and so on (see Figure 1). Both, the number of families per population and the number of individuals per family can vary, but the number of landmarks per individual should be constant and higher than three throughout the entire data set. Example input files for both kind of data are available at the online help in the program.

Figure 1.

Morphometric analysis

In the study of form differences, three different components can be considered: centroid size, uniform (affine) and non-uniform (non-affine) shape differences (Rohlf et al., 1996 and references therein). Centroid size is used as the preferred measure of the geometric size of each individual because of its properties of independence with respect to shape variables, when landmark scatter meets the assumption of circular landmark location errors with the same variance at all landmarks (the null model), and there is no allometry. Thus, allometry can be tested via multiple regression of centroid size upon shape variables (Bookstein, 1991). The uniform component of shape describes changes that affect to the whole set of landmarks. In short, the non-uniform component of shape variation describes local deformations in shape.

In MODICOS, after choosing the morphometric analysis option, three options are displayed (Figure 2): Distance and Centroid Size, RWA (Relative Warp Analysis) and Uniform.

Distance and Centroid Size

For each individual, the program computes the centroid size and the Euclidean distances between landmarks. Centroid size is computed as the square root of the sum of squared distances of a set of landmarks from their centroid. The output is written in a text file so that rows are individuals, the first column is the individual identifier, the following columns are the distance values, and the last two columns are the centroid and the centroid size. Each individual centroid size remains in memory, so that it can be immediately used as a phenotype for the genetic analysis.

RWA (Relative Warp Analysis)

Consider the multidimensional non-Euclidean space of all possible shapes of configurations of landmark points, this shape space is known as Kendall's shape space. When variation in shape is sufficiently small, Kendall's space can be well-approximated by a Euclidean space tangent to it (Rohlf and Bookstein, 2003 and references therein). This Euclidean space can be decomposed into the direct sum of the uniform and the non-uniform subspaces. Therefore, any given set of p landmarks representing a shape can be expressed as a vector s with $2p - 4$ elements so that

$$s = u + b \qquad (1)$$

Any vector s of the tangent space (plane) is a sum of a vector u from the uniform subspace and other vector b for the non-uniform subspace. Thus, given a shape, if we have the uniform component we can compute the non-uniform one and *viceversa* (Rohlf and Bookstein, 2003).

RWA, Relative Warp Analysis uses thin-plate spline technique (Bookstein, 1991; Rohlf, 1993) to capture the non-uniform shape variation. The relative warps are principal components of the distribution of shapes in the tangent space and they are computed following the algorithm given in Rohlf (1993, 1998). For each individual we obtain a relative warp vector of dimension min {2(p -

3), *n* -1}, where *n* is the number of measured objects. Thus, the major axes of variation are captured as relative warps which provide orthogonal descriptors of some fraction of the variation.

Uniform

The uniform shape component is computed using the complement of the space of pure bending shape variation (Rohlf and Bookstein, 2003). For each individual, a vector of two uncorrelated components is obtained.

The results from the RWA and Uniform analyses are written into a file in a format similar to that of the Euclidean distances and centroid size (see above). Again, results are maintained in memory so they can be used in a genetic analysis. In the case of RWA two files, "RC.dat" and "R.dat", are also written with the reference configuration and relative warp loadings matrices respectively (see Rohlf, 1993 for an explanation on relative warp loadings and graphical representation of shape variation). These matrices allow for the graphical representation of the importance of each landmark for the deviation from the reference configuration of the sample.

Figure 2.

Genetic analysis

Geometric morphometric can be combined with quantitative genetic methods to estimate components of genetic variation for quantitative traits such as size and shape (Palsson and Gibson, 2004). After pushing the genetic analysis button, two options are displayed: "Halfsib" and "Fullsib", referring to the family design used for the estimation of components of genetic variation. MODICOS makes a population partition of morphometric measurements (phenotypes) in terms of

quantitative genetics. In quantitative traits the total variance (phenotypic variance, V_P) is partitioned as follows

$V_P = V_G + V_E$ (2)

where V_G is the genetic variance and V_E is the environmental variance. The genetic variance is also partitioned in additive, dominance and interaction components. The phenotypic resemblance between relatives (variance between families, θ^2_{fam}) can be used to estimate the additive genetic variance components (V_A) using both, fullsib ($\theta^2_{fam} \cong V_A/2$) or halfsib ($\theta^2_{fam} = V_A/4$) designs (Falconer and Mackay, 1996). This additive genetic variance computation is performed within each population via one-factor ANOVA since the between-family variance component θ^2_{fam} is equal to the covariance of the family members (Falconer and Mackay, 1996). There are other possible more complex designs but field data usually comes in these two fashions.

In MODICOS, centroid size or any relative warp or uniform component can be used as a phenotype to compute V_A estimates. The estimated V_A variance (averaged across populations) is also the within-population variance component, and jointly with the between component (V_b), allows for the Q_{ST} computation in a two factor (population, family) nested ANOVA (Spitze, 1993) as

$Q_{ST} = V_b / (V_b + 2V_A)$ (3).

VALIDATION AND RESULTS

To check the morphometric computations in MODICOS some freely available software was used: Morpheus (Slice, 1993) was used to check the Procrustes Superimposition (necessary for shape analyses) and tpsRelw (Rohlf, 1998) to check procedures of shape analysis (relative warp and uniform procedures). In all cases the output of MODICOS was almost identical (up to 4th or 5th

decimal position) to that of the programs above. Genetic computations as ANOVAS, heritability and *QST* were checked both by hand and using standard statistical packages.

To show the use of MODICOS, a complete analysis (morphometric and genetic) was performed using a sample of two different ecotypes of *Littorina saxatilis* (rough periwinkle), each with 29 and 30 families respectively, and 3 individuals per family. Previous work has shown significant genetic variability at different shore levels for shell height (Carballo *et al*., 2001). Now we considered shell size using centroid size and shell shape via the global component i.e. the two (*x* and *y*) uniform shape phenotypes and the local component (the relative warps). Each individual consists of a configuration of 12 landmarks (Figure 3). Images of individuals were digitized using an image analyzer and a binocular. For the genetic analysis we compared the inter-population genetic variance estimates for the centroid size phenotype, the two uniform shape phenotypes and the two first relative warps. The selected design for the genetic analysis was the full-sib one.

Figure 3

The inter-population genetic analysis gave significant differences for centroid size, the first uniform component and the two first relative warps (Table 1). Interestingly, the Q_{ST} estimate is higher (as far as twice) for the relative warps than for size and the uniform component. Since allometry was no detected, neither for the uniform component nor for the two relative warps, the higher Q_{ST} value seems to indicate that the non-uniform shape component is more efficient capturing the effect of natural selection onto the two different ecotypes. This kind of effect is better visualized in a graphical way using the scaled scores for the two first relative warps providing an

ordination space (Rohlf 1993). For an example of the later see Figures 3 and 4 in Carvajal-Rodríguez *et al* (2005a).

Table 1.

CONCLUSION

MODICOS is a computer program with user-friendly Windows 95/98/2000/NT/XP interface. It has a friendly on-line help interface with detailed information on input formats, output, methods and program features. By running MODICOS users can perform morphometric analysis onto a data set, getting for each individual one quantitative trait phenotype as is centroid size, two more phenotypes as uniform components of shape variation and, in the case of, say, using 12 landmarks, up to 18 phenotypes of local components of shape variation. All these phenotypes can be immediately used to estimate heritable genetic variance components within and between populations. As shown via the example above, it could be of interest to consider shape separated from size in inter-population quantitative genetic analysis.

REFERENCES

Arqvinst, G. and Thornhill, R. (1998). Evolution of animal genitalia: patterns of phenotypic and genotypic variation and condition dependence of genital and non-genital morphology in water strider (Heteroptera: Gerridae: Insecta). Genet. Res., 71, 193-212.

Bookstein, FL. (1991). Morphometric tools for landmark data. Cambridge University Press.

Carballo, M., Garcia, C. and Rolán-Alvarez, E. (2001). Heritability of shell traits in wild *Littorina saxatilis* populations: results across a hybrid zone. Journal of shellfish research, 20, 415-422.

Carvajal-Rodríguez, A., Conde-Padín, P. and Rolán-Alvarez, E. (2005a). Decomposing shell form into size and shape by geometric morphometric methods in two sympatric ecotypes of *Littorina saxatilis*. Journal of Molluscan Studies. In Press.

Carvajal-Rodríguez, A., Rolán-Álvarez, E. and Caballero, A. (2005b). Quantitative variation as a tool for detecting human-induced impacts on genetic diversity. Biological Conservation. In Press.

Cavalcanti, MJ., Monteiro, RL. and Lopes, PRD. (1999). Landmark-based morphometric analysis in selected species of serranid fishes (Perciformes: Teleostei). Zoological Studies, 38, 287-294.

Douglas, ME., Douglas, MR., Lynch, JM. and McElroy, DM. (2001). Use of geometric morphometrics to differentiate Gila (Cyprinidae) within the upper Colorado river basin. Copeia, 2, 389-400.

Falconer, DS. and MacKay, TFC. (1996). *Introduction to quantitative genetics.* Longman, Essex.

Klingenberg, CP., Leamy, LJ., Routman, EJ. and Cheverud, JM. (2001). Genetic architecture of mandible shape in mice: effects of quantitative trait loci analyzed by geometric morphometrics. Genetics, 157, 785-802.

McKay, J.K. and Latta, R.G. (2002). Adaptive population divergence: markers, QTL and traits. Trends Ecol. Evol., 17, 285-291.

Monteiro, RL., Bordin, B. and Furtado dos Reis, S. (2000). Shape distances, shape spaces and the comparison of morphometric methods. TREE, 15(6), 217-220.

Palsson, A. and Gibson, G. (2004). Association between nucleotide variation in Egfr and wing shape in *Drosophila melanogaster*. Genetics, 167, 1187-1198.

Rohlf, FJ. (1993). Relative warp analysis and an example of its application to mosquito wings. Pages 131-159 in Contributions to morphometrics, volume 8 (L.F Marcus, E. Bello and A. García Valdecasas, eds.). Museo Nacional de Ciencias Naturales, Madrid.

Rohlf, FJ. (1998). TPSrelw: relative warps, version 1.20. N.Y.: State Univ. at Stony Brook.

Rohlf, FJ., Loy, A. and Corti, M. (1996). Morphometric analysis of old world Talpidae (Mammalia, Insectivora) using partial-warp scores. Systematic Biology, 45(3), 344-362.

Rohlf, FJ. and Bookstein, FL. (2003). Computing the uniform component of shape variation. Systematic Biology, 52(1), 66-69.

Rohlf, FJ. and Marcus, LF. (1993). A revolution in morphometrics. TREE, 8, 129-132.

Rohlf, FJ. and Slice, D. (1990). Extensions of the Procustes method for the optimal superimposition of landmarks. Syst. Zool, 39, 40-59.

Slice, D. (1993). Morpheus *et al*. N.Y.: State University at Stony Brook.

Spitze, K. (1993). Population structure in *Daphnia obtusa*: Quantitative genetic and allozymic variation. Genetics, 135, 367-374.

Wright, S. (1951). The genetic structure of populations. Annu. Eugenics, 15, 323-354.

Acknowledgements

We wish to thanks Emilio Rolán-Alvarez for comments on the manuscript and also the original idea from which MODICOS arise. Professor Rohlf F.J provided very useful comments for the debugging work of Procrustes superimposition and RWA algorithm implementations. Conde-Padín kindly gave her data for the example. Thanks also to D. Posada for English corrections. Some freely available software was used to check morphometric computations in MODICOS: Morpheus (Slice, 1993) and tpsRelw (Rohlf, 1998).

This work is part of Eumar european biodiversity project (EUK3-CT-2001-00048).

CAPTION TO FIGURES

Figure 1. Morphometric data file format used in MODICOS: The first individual identifier, 111.01.0, represents population 111, family 1 and individual 0; 111.01.1 is the individual 1 from same population and family. The pairs separated by commas under each individual identifier are the coordinate points of each individual configuration.

Figure 2. Screen shots of Morphometric analysis window.

Figure 3. Distribution of landmark points for the morphometric analysis used in *Littorina saxatilis*.

FIGURES

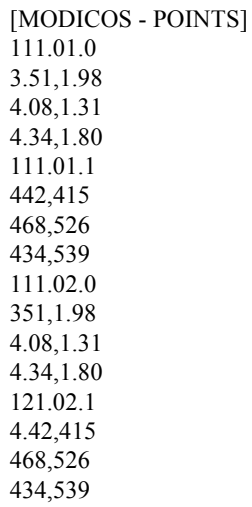

Figure 1

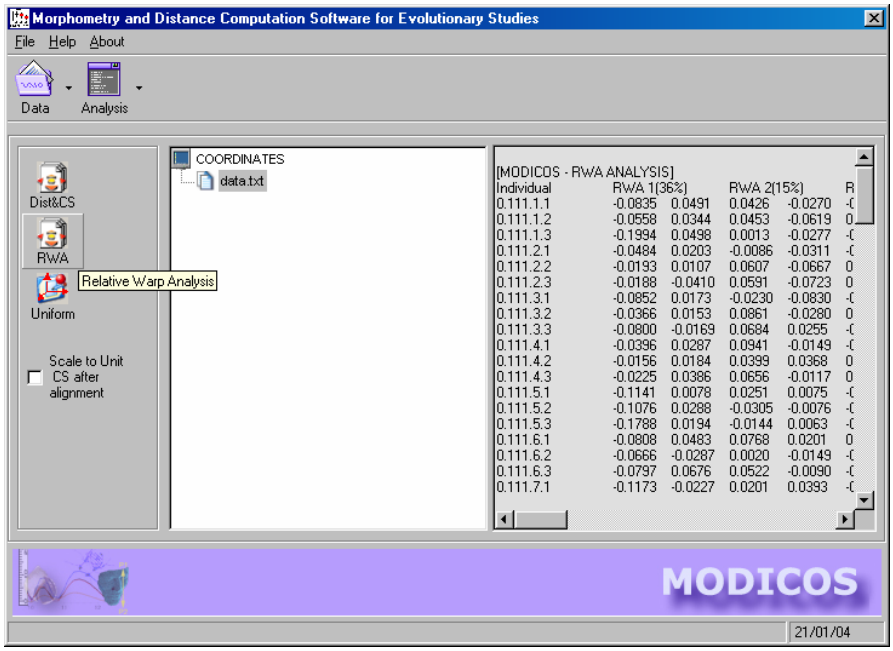
Morphometry and Distance Computation Software for Evolutionary Studies
File Help About
Data
Analysis
Dist&CS
RWA
Relative Warp Analysis
Uniform
Scale to Unit CS after alignment
COORDINATES
data.txt
[MODICOS - RWA ANALYSIS]
Individual RWA 1(36%) RWA 2(15%)
0.111.1.1 -0.0835 0.0491 0.0426 -0.0270
0.111.1.2 -0.0558 0.0344 0.0453 -0.0619
0.111.1.3 -0.1994 0.0498 0.0013 -0.0277
0.111.2.1 -0.0484 0.0203 -0.0086 -0.0311
0.111.2.2 -0.0193 0.0107 0.0607 -0.0667
0.111.2.3 -0.0188 -0.0410 0.0591 -0.0723
0.111.3.1 -0.0852 0.0173 -0.0230 -0.0830
0.111.3.2 -0.0366 0.0153 0.0861 -0.0280
0.111.3.3 -0.0800 -0.0169 0.0684 0.0255
0.111.4.1 -0.0396 0.0287 0.0941 -0.0149
0.111.4.2 -0.0156 0.0184 0.0399 0.0368
0.111.4.3 -0.0225 0.0386 0.0656 -0.0117
0.111.5.1 -0.1141 0.0078 0.0251 0.0075
0.111.5.2 -0.1076 0.0288 -0.0305 -0.0076
0.111.5.3 -0.1788 0.0194 -0.0144 0.0063
0.111.6.1 -0.0808 0.0483 0.0768 0.0201
0.111.6.2 -0.0666 -0.0287 0.0020 -0.0149
0.111.6.3 -0.0797 0.0676 0.0522 -0.0090
0.111.7.1 -0.1173 -0.0227 0.0201 0.0393
MODICOS
21/01/04

Figure 2

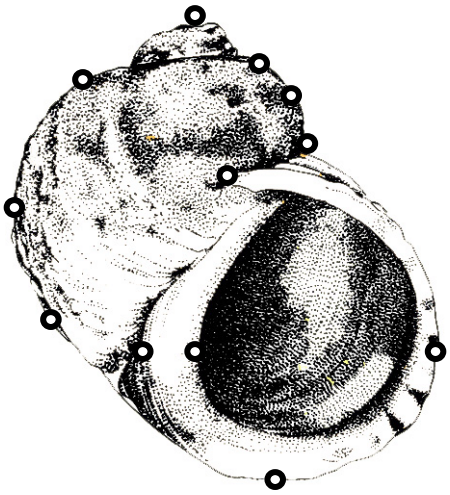

Figure 3

TABLE

Table 1. Result of two factor nested ANOVA in Q_{ST} computation for the phenotypes: centroid size (CS), uniform components of shape (U1 and U2) and first two relative warps (RW1 and RW2).

Phenotype	**F among populations**	***QST***
CS	30.2 ***	0.26
U1	24.8 ***	0.26
U2	0.17 ns	0.0
RW1	66.5 ***	0.60
RW2	63.4 ***	0.53