A new application of morphometrics in a study of the variation in uncinal shape present within the Terebellidae (Polychaeta): a reevaluation from digital images

André Rinaldo Senna GARRAFFONI[1]* and Maurício de Garcia CAMARGO[2]

[1] *Universidade Estadual de Campinas, Instituto de Biologia, Departamento de Zoologia, R. Charles Darwin s/n, CP 6109, Campinas, SP, Brazil,13083-970. *Corresponding author: email andregarraffoni@yahoo.com.br*

[2] *Centro de Estudos do Mar, Universidade Federal do Paraná, Av. Beira mar s/n, CP 50002, Pontal do Paraná, PR, Brazil, 83255-000*

Abstract: In this study, the morphometric approach was used to establish distinct morphological groups in regard to uncini shape within the four subfamilies of Terebellidae (Polychaeta). To achieve this objective, 24 distances were measured, based on photographs of the three uncini dissected from segment 7 and another three on segment 16, from 31 species, 2 belonging to Trichobranchinae, 2 to Polycirrinae, 6 to Thelepodinae and 21 to Terebellinae. Those distances were based on 7 real landmarks and 6 "extrapolated" landmarks, and measurements of the area, perimeter, greatest length, shortest length, diagonal perimeter, horizontal perimeter and vertical perimeter. The multidimensional approach to assess the similarity and dissimilarity of the distinct uncini patterns was done using n-MDS, ANOSIM and PCA. A prediction model was developed to identify the Terebellidae subfamily based on neural network. The n-MDS performed on the 31 Terebellidae species recognized three distinct groups: one group composed of species belonging to the subfamilies Thelepodinae-Terebellinae and two other distinct groups formed by the species of Trichobranchinae and Polycirrinae. The permutation tested by ANOSIM confirmed the trends observed in the n-MDS. The ordination of the two PCA axes explained 64.20% of the variance on the PCA realized on uncini of segment 7 and 65.05% of the variance on the PCA realized on uncini of segment 16. The PCA showed that most of the measurements made on the uncini were important to obtain a better classification of the data. Thus, these three different patterns can be assumed to be three different character states that define the overall uncini shape and can be used in future cladistics analyses. Finally, the model developed based on the neural network showed good success in classifying the four subfamilies.

Résumé : *Une nouvelle application de la morphométrie à l'étude de la variation de forme des uncini présents chez les Terebellidae (Polychaeta) : une réévaluation à partir d'images numériques.* Dans cette étude, l'approche morphométriques est utilisée pour définir des groupes distincts en fonction de la forme des uncini dans quatre sous-familles de Terebellidae (Polychaeta). 24 paramètres ont été mesurés à partir de photographies de 3 uncini issues du segment 7 et de 3 autres du segment 16, sur 31 espèces, 2 appartenant aux Trichobranchinae, 2 aux Polycirrinae, 6 aux Thelepodinae et 21 aux Terebellinae. Ces paramètres sont fondés sur 7 limites précises et 6 limites extrapolées ainsi que la mesure de la surface, du périmètre, de la plus grande longueur, de la plus petite longueur, du périmètre diagonal, du périmètre horizontal et du périmètre vertical. L'approche multidimensionnelle pour estimer la similarité et la dissimilarité des différents patrons d'uncini a été réalisée grâce à des n-MDS, ANOSIM et ACP. Le modèle prédictif développé pour identifier les sous-familles

de Terebellidae est fondé sur l'approche par réseau de neurones. Le n-MDS réalisé sur les 31 espèces de Terebellidae met en évidence 3 groupes distincts : un groupe composé d'espèces appartenant aux sous-familles Thelepodinae-Terebellinae et deux autres groupes formés par les espèces de Trichobranchinae et Polycirrinae. La permutation testée par ANOSIM a confirmé les tendances observées dans le n-MDS. L'ordination sur les deux premiers axes de l'ACP explique respective-ment 64,2% de la variance totale sur l'ACP réalisée sur les uncini du segment 7 et 65,05% sur l'ACP réalisée sur les uncini du segment 16. Les ACP montrent que la majorité des mesures effectuées sur les uncini est importante pour obtenir une meilleure classification des données. Les trois patrons obtenus peuvent donc être considérés comme trois états caractéris-tiques pour définir la forme globale des uncini et ainsi utilisés dans de futures études cladistiques. Finalement, le modèle développé sur la base d'une réseau de neurones a permis de séparer avec succès les quatre sous-familles.

Keywords: Coding ● Terebelliformia ● Artificial Neural Network ● Neurochaetae ● Multivariate analysis.

Introduction

Recent taxonomic studies have used morphometric approaches to detect and characterize possible new features that could help to improve the descriptions both of new species that are phylogenetically closely related and of new characters for cladistics analysis (Chimimba et al., 1999; Roth & Mercer, 2000; Jordaens et al., 2002; Guerrero et al., 2003; Garraffoni & Camargo, 2006; Glasby & Glasby, 2006). This kind of approach allows a rigorous quantitative analysis of variation in organism size and shape for morphometric purposes, based on multivariate statistical analysis (Kligenberg, 2002). Refinement of the quantitative description of variation helps to assess intermediate shapes, judge the degree of similarity and proximity, and extrapo-late or predict hypotheses from experimental or nonexistent extremes (Roth & Mercer, 2000). The overall shape of an organism or a structure and the variability among related groups are the basis of understanding its evolution and the interaction among the phenotype, genotype and environ-ment (Monteiro et al., 2002).

There are many approaches dealing with morphometrics including traditional and geometric ones. A very common approach using morphometrics is to consider the land-marks, the configuration of morphological features (Kligenberg, 2002). Landmarks can be interpreted as a way to reduce the shape of an organism to a set of points that characterize its general traits, which are assumed to be homologous across individuals and populations (Rolf & Marcus, 1993; David & Laurin, 1998; Costa et al., 2004) and in some cases to recognize species (Becerra & Valdecasa, 2004). Landmarks can be defined operationally as relocatable coordinate positions on an object in a two-dimensional or three-dimensional Euclidean measurement space (MacLeod, 2001) using a set of measurements (variables) between points established *a* priori on the analysed structure.

The use of morphometry as a tool to help to define distinct patterns within one feature, to split or lump different "species," or to define quantitative features within the Polychaeta is in the beginning stages. Ben-Eliahu (1987) and Fauchald (1991) made the first forays into morphometric studies when they established and compared some measurements, respectively, from the notopodium of some nereidids and the total length of some eunicid species, without any statistical support. Orrhage & Sundberg (1990) were the first to use statistical analyses, such as principal components analysis and canonical analysis, to show the presence of three different morphological patterns within the *Laonice* complex. Recently, Maltagliati et al. (2001), Garraffoni & Camargo (2006), Glasby & Glasby (2006), and Costa-Paiva & Paiva (2007) performed morphometric studies using other statistical tools, including cluster analysis, Principal Components Analysis (Relative Warps Analysis), and n-MDS (non-metric Multidimensional Scaling ordination).

Garraffoni & Camargo (2006) used landmark methodology applied to the uncini of terebellids, in attempting to elucidate the overall performance of morphological variations in the uncini, and also to assess the possible taxonomic information available. Garraffoni & Camargo (2006) stated that three different shape components showed significant differences among the four subfamilies. These three different shapes can also be taken as three character states for the character overall uncinal shape. Although the Trichobranchinae and Polycirrinae have their own overall shape, Terebellinae and Thelepodinae morphotypes are very similar, and it is diffi-cult to divide them into different states. Garraffoni & Camargo (2006) also pointed out that the posterior process

present in some species of Terebellinae is not homologous with the long shaft manubrium present in the species of Trichobranchinae.

The present study aimed to improve on the studies initiated by Garraffoni & Camargo (2006), using distances measured directly from the uncini dissected from terebellid specimens, and analysing these new data using other multivariate analyses that help to refine the characterization of the different morphological patterns. We also developed a computer program based on an artificial neural network that allows predictions about the Terebellidae subfamily classification, using the measurements of the uncini.

Materials and Methods

Terebellidae species

A total of 62 specimens were analysed, representing 31 species of Terebellidae: 2 species belonging to Trichobranchinae (6 specimens), 2 species belonging to Polycirrinae (2 specimens), 6 species belonging to Thelepodinae (7 specimens), and 21 species belonging to Terebellinae (47 specimens) (Table 1). The predominance of species of Terebellinae used in the present study reflects the greater diversity of the morphological characters (including the variability in uncini shape) in this sub-group compared with the other family members, as noted previously by Garraffoni & Camargo (2006).

For the present study, we analysed different species from those studied by Garraffoni & Camargo (2006), because of the difficult in obtaining specimens to dissect the neuropodium. However, we attempted to analyse the same number of species and genera from each subfamily as in the preview study. We also analysed two policirrins species that fitted in each of the distinct types of uncini identified by Glasby & Glasby (2006).

Measurements and morphometric identifications

The measurements employed in the present analysis express modifications in structures inferred to be homologous within the species (Reis, 1988). Three uncini were dissected from neuropodia present on segment 7 and another three from neuropodia present on segment 16 of each of the 62 specimens analysed. These two segments were chosen to exemplify the anterior and posterior segments of the anterior region of the body (or thoracic segment), in order to better characterize possible morphological variation in the uncini of each region, as well as the variability within each neuropodium. The species *Laphania boecki¸ Polycirrus aurantiacus*, and *P. paivai* do not have uncini on segment 7, and for this reason we dissected the uncini from the first segment bearing

Table 1. List of terebelid species used in the present analysis.
Tableau 1. Liste des espèces de Térébellidés étudiés.

Species name	Subfamily	Abbreviation
Polycirrus aurantiacus	Polycirrinae	Po_au
Polycirrus paivai	Polycirrinae	Po_pa
Amphitrite cirrata	Terebellinae	Amp_ci
Amphitritides gracilis	Terebellinae	Amph_gr
Artacama benedeni	Terebellinae	Art_be
Axionice maculata	Terebellinae	Ax_ma
Eupolymnia congruens	Terebellinae	Eu_con
Eupolymnia sp.	Terebellinae	Eu_eu
Lanice conchilega	Terebellinae	La_co
Laphania boecki	Terebellinae	Lap_bo
Loimia sp. 1	Terebellinae	Lo_jp
Loimia sp. 2	Terebellinae	Lo_pgua
Neamphitrite variabilis	Terebellinae	Ne_va
Neoamphitrite figulus	Terebellinae	Neo_fi
Nicolea gracilibranchus	Terebellinae	Ni_gra
Nicolea venustula	Terebellinae	Ni_ve
Paraeupolymnia uspiana	Terebellinae	Pa_usp
Pista corrientis	Terebellinae	Pi_co
Pista elongata	Terebellinae	Pi_elo
Pista sp. 1	Terebellinae	Pi_jp
Pista sp. 2	Terebellinae	Pi_pgua
Proclea malmgreni	Terebellinae	Pro_mal
Terebella lapidaria	Terebellinae	Te_lap
Parathelepus collaris	Thelepodinae	Para_co
Streblosoma hartmanae	Thelepodinae	Ste_ha
Streblosoma sp. 1	Thelepodinae	Ste_seb
Thelepus cincinnatus	Thelepodinae	The_ci
Thelepus sp. 1	Thelepodinae	The_chi
Thelepus sp. 2	Thelepodinae	The_pgua
Terebellides anguicomus	Trichobranchinae	Tere_an
Terebellides kobei	Trichobranchinae	Tere_ko

uncini, i.e., segments 9, 12, and 11, respectively. The uncini were photographed and measured by the computer program Jandel SigmaScan® Pro 2.0.

In the present analysis, 17 distances were measured using 7 real landmarks (landmarks which can be defined as homologous anatomical loci among the different chaeta) and another 6 points that are extrapolations of the real landmarks (landmarks that not fall directly on the analysed structure) to show the differences observed in the morphology of the thoracic uncini in Terebellidae (Table 2, Fig. 1). A computer program allowed us to measure 7 distances not studied by Garraffoni & Camargo (2006) as area (distance 18), perimeter (distance 19), greatest length (distance 20), shortest length (distance 21), diagonal perimeter (distance 22), horizontal perimeter (distance 23), and vertical perimeter (distance 24).

Bookstein (1991) suggested that three main kinds of landmarks may be used: juxtaposition, maximum curvature, and external constructed points, and also intermediary
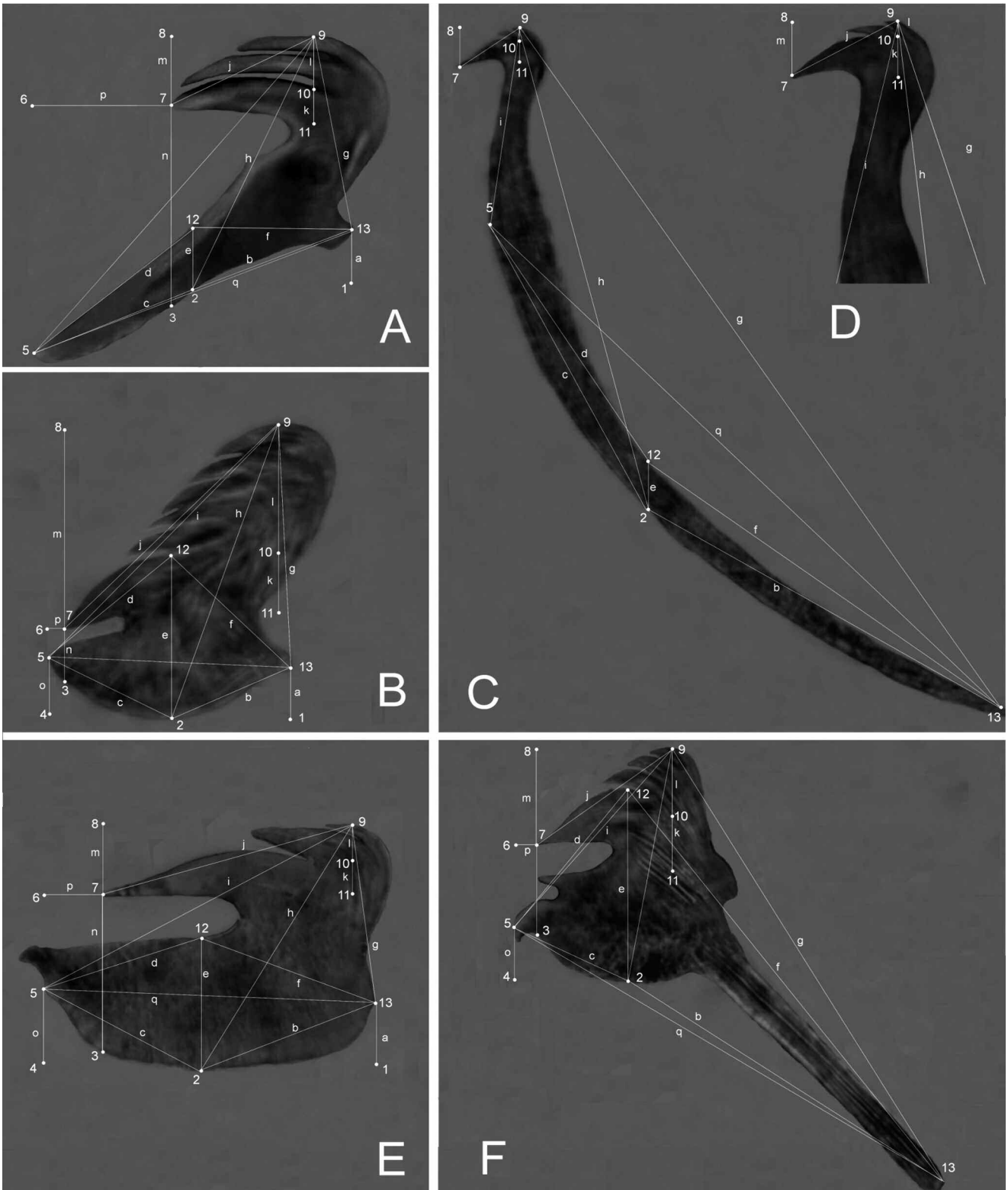
Figure 1. Morphometric measurements taken on uncini of the Terebellidae species. **A.** *Polycirrus aurantiacus,* Polycirrinae. **B.** *Artacama benedeni*, Terebellinae. **C-D.** *Terebellides anguicomus*, Trichobranchinae. **E.** *Thelepus* sp. 1, Thelepodinae. **F.** *Pista* sp., Terebellinae.

Figure 1. Mesures morphométriques prises sur les uncini des Terebellidae. **A.** *Polycirrus aurantiacus,* Polycirrinae. **B.** *Artacama benedeni*, Terebellinae. **C-D.** *Terebellides anguicomus*, Trichobranchinae. **E.** *Thelepus* sp. 1, Thelepodinae. **F.** *Pista* sp., Terebellinae.

Table 2. Distances used in the present analysis.
Tableau 2. Distances utilisées dans cette étude.

Distances	Initial landmark	Final landmark
1 (e)	12	2
2 (q)	13	5
3 (l)	9	10
4 (k)	9	11
5 (g)	13	9
6 (h)	9	2
7 (i)	9	5
8 (j)	9	7
9 (a)	1	13
10 (o)	5	4
11 (n)	7	3
12 (p)	7	6
13 (m)	8	7
14 (b)	13	2
15 (c)	2	5
16 (d)	5	12
17 (f)	12	13

cases between types 2 and 3. We used an intermediary case between types 2 and 3 (David & Laurin, 1998), because of the nature of our data (distances were measured between different landmarks arranged in an irregularly shaped structure).

Statistical Analysis

The use of the multivariate statistic is essential to a consistent analysis of the measurements obtained in the meristic variables or characteristics in a morphometric study (Monteiro & Reis, 1999). The proximity multivariate analyses (n-MDS), ANOSIM, and principal component analysis (PCA), plus the artificial neural network (ANN) were used.

While the classical multivariate analyses (n-MDS, ANOSIM, and PCA) only perceive linear relations among groups (however, n-MDS can show groups from a similarity matrix in a non-metric way), the neural network allows one to distinguish non-linear groups, because of its ability of self-learning using examples and generalizations obtained from past information.

The analytical steps used in the present study were the same used by Garraffoni & Camargo (2006). The measurements were standardized, transformed by log (X+1), and the Euclidean distance was applied. The software Primer version 5.2.4 was used to carry out the n-MDS analysis. This analysis gives a general overview of the data and can be used for illuminating relationships among species. The n-MDS can construct a map or configuration of the objects (Clarke & Warwick, 1994). The graphic distortions produced by the necessity to regulate the distances are

represented by a stress factor that shows the fidelity of the data plotted on the graph, in relation to the original similarity matrix. The ANOSIM (Analysis of Similarities) is a permutation test analogous to the standard ANOVA (Analysis of Variance), used to detect differences among groups established *a priori* in n-MDS (Clarke & Warwick, 1994). ANOSIM was used here to test possible groups of species belonging to the same subfamilies (Terebellidae, Trichobranchinae, Polycirrinae, and Thelepodinae) according to the ordination.

Principal Components Analysis

Principal components analysis (PCA) was used because it does not depend on an *a priori* classification of the species, and quantifies the different shapes independently. The PCA is an exploratory method, because it produces hypothesis from the data obtained in the study, and does not use any hypothesis inferred from another study. PCA represents a way of projecting points from the original, high-dimensional variable space onto a two-dimensional plane, with minimal loss of information. The analysis of the PCA was done using the software STATISTICA version 6.0.

Artificial neural network

An artificial neural network (ANN) is a computational model that consists of processing elements (i.e., nodes or neurons) and connections between them, as well as training and recalling algorithms. This model is analogous to the neural structure from biological organisms that have neural systems comprised of billions of neuron connections. The neural network is not based on *a priori* rules, but on learning and training patterns presented in the network.

The intelligent behavior of the ANN originates from the interactions among the units in the network procedures. There are several procedures, the most popular of which is known as 'back-propagation' (Rumelhart et al., 1986), for adjusting the connection weights appropriately during training. Back-propagation is essentially a neural network implementation of gradient descent, and requires the specification of some additional parameters, e.g., the learning rate. The aim of this procedure is to find the connections that minimize the prediction error of the network.

A neural network typically comprises three types of layers: an input layer, the hidden or intermediate layers, and an output layer (Olden, 2000; Medeiros, 2003; Dedecker et al., 2004; Tatibana & Kaetsu, 2005). The input layer provides the corresponding value that will be classified during the analysis (in the present study, 17 neurons were used in the input layer, representing the 17 distances measured in the uncini). The complexity of the neural network will be directly related to the number of neurons in the input layer. The final number of neurons present in

input layer must be the same number of neurons present in the hidden layer. In the output layer we found the neurons responsible for the conclusion and presentation of the final results (four neurons representing the four valid Terebellidae subfamilies used in the present study).

The first step to create an ANN is the collection of the data for a specific problem (in this study, the measurements obtained from different Terebellidae uncini) to establish a set of data that will be used for the training (the data matrix with the measures), and then to test the model (Medeiros, 2003; Carvalho, 2005; Tatibana & Kaetsu, 2005).

The construction and implementation of the ANN developed in the present paper was done using the software Delphi 6 with the component TMLP developed by Medeiros (2003). The data matrix, as well as the software developed in the present paper, is available at http://www.cem.ufpr.br/garraffoni/redeneural.

Results

Before we implemented the multivariate analysis and the artificial neural network, we tested the colinearity among the original data measurements, because this test could show whether we were using distances that were redundant (auto-correlation). Thus, after this test, we removed the distances 3, 11, 12, 21, 22, 23, and 24 from the data matrix for segments 7 and 16.

The result of the n-MDS applied to the uncini from segments 7 and 16 of the 31 terebellid species (Figs. 2 & 3) indicated a very low stress value (2-d: Minimum stress, 0.1 on segment 7; 2-d: Minimum stress, 0.11 on segment 16), demonstrating high resolution (Clarke & Warwick, 1994; Chui et al., 2002). The n-MDS analysis using 17 characters clearly separated the species into 3 clusters according to different quadrants (Fig. 2). Trichobranchinae was clustered in the 1st quadrant, Polycirrinae in the 2nd quadrant, and

Thelepodinae and Terebellinae in the 3rd quadrant. Only the thelepodine species *Parathelepus colaris* was not grouped with other thelepodine-terebelline species.

The permutation test ANOSIM (Table 3) confirmed the tendency observed in the n-MDS, because each pair of groups was significantly separated, except Terebellidae against the Thelepodinae. In this test the R statistic is very important to aid in discriminating which pair is united or distinct over a range that begins at 0 (low power of discrimination) and ends at 1 (high power of discrimination). In the present study, almost all combinations among the groups Terebellinae, Trichobranchinae, Thelepodinae and Polycirrinae (Tri-The, Tri-Poly, Tri-Tere, Tere-Poly) showed values close to 1 and only the combination (Thel-Poly) showed values close to 0.5. However, as pointed out above, the pair Thelepodinae-Terebellinae was the exception and showed a value close to 0, showing a low power of discrimination.

The PCA showed that the first two principal components distances measured in the uncini from segment 7 accounted for 64.20% of the total variance (34.64% for PC1, 29.55% for PC2). In the first component, almost all the distances showed positive correlations; only the distances 17, 18, and 19 showed negative correlations (Tables 4 & 5, Fig. 4). In the second component, the distances 4, 9, 10, 15, 16, 17, 18, 19, and 20 showed positive correlations; the distances 1, 2, 5, 6, 7, 13, and 14 showed negative correlations.

The principal components analysis showed that the first two principal components distances measured in the uncini from segment 16 accounted for 65.06% of the total variance (32.78% for PC1, 32.27% for PC2). In the first component, almost all the distances showed positive correlations; only the distances 9, 17, 18, 19, and 20 showed negative correlations (Tables 4 & 5, Fig. 5). In the second component, almost all the distances showed positive correlations; only the distances 7, 13, and 14 showed negative correlations.

Table 3. Results of the ANOSIM (Global significance = 0.1%, number of permutations = 999) of the uncini on segment 7 (Global R = 0.65) and segment 16 (Global R = 0.62), from differents Terebellidae subfamilies. Abbreviations: Tere- Terebellinae, Tricho- Trichobranchinae, The- Thelepodinae, Poly- Polycirrinae.

Tableau 3. Résultats de L'ANOSIM (significativité globale = 0,1%, nombre de permutations = 999) réalisée sur les uncini du segment 7 (R global = 0,65) et du segment 16 (R global = 0,62), des différentes sous-familles de Terebellidae. Abréviations : Tere- Terebellinae, Tricho- Trichobranchinae, The- Thelepodinae, Poly- Polycirrinae.

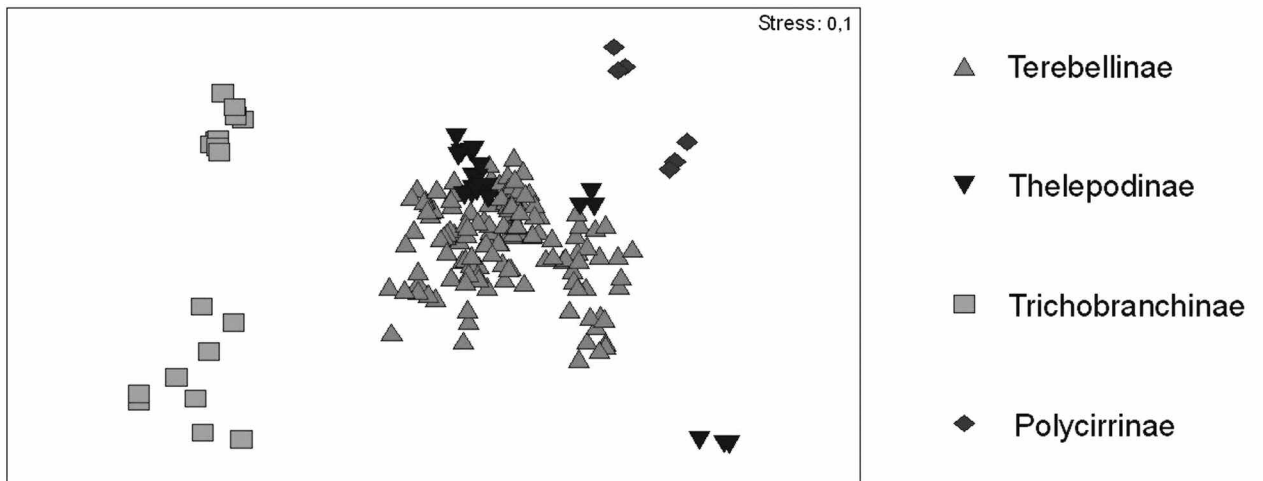| | Segment 7 | | | | Segment 16 | | |
Groups	R Statistic	Significance Level (p)	Possible Permutations	Groups	R statistic	Significance Level (p)	Possible Permutations
Tere, Thel	0.206	0.3	Many	Tere, Thel	0.166	1.5	Many
Tere, Tri	0.991	0.1	Many	Tere, Tri	0.990	0.1	Many
Tere, Poly	0.894	0.1	Many	Tere, Poly	0.862	0.1	Many
Thel, Tri	0.699	0.1	Many	Thel, Tri	0.688	0.1	Many
Thel, Poly	0.524	0.2	296010	Thel, Poly	0.507	0.3	296010
Tri, Poly	0.999	0.1	134596	Tri, Poly	0.938	0.1	134596

Figure 2. Two-dimensional plot of 17 measurements on uncini from segment 7 analyzed by non-metric multidimensional scaling.
Figure 2. Représentation bidimensionnelle par n-MDS des 17 mesures réalisées sur les uncini du segment 7.
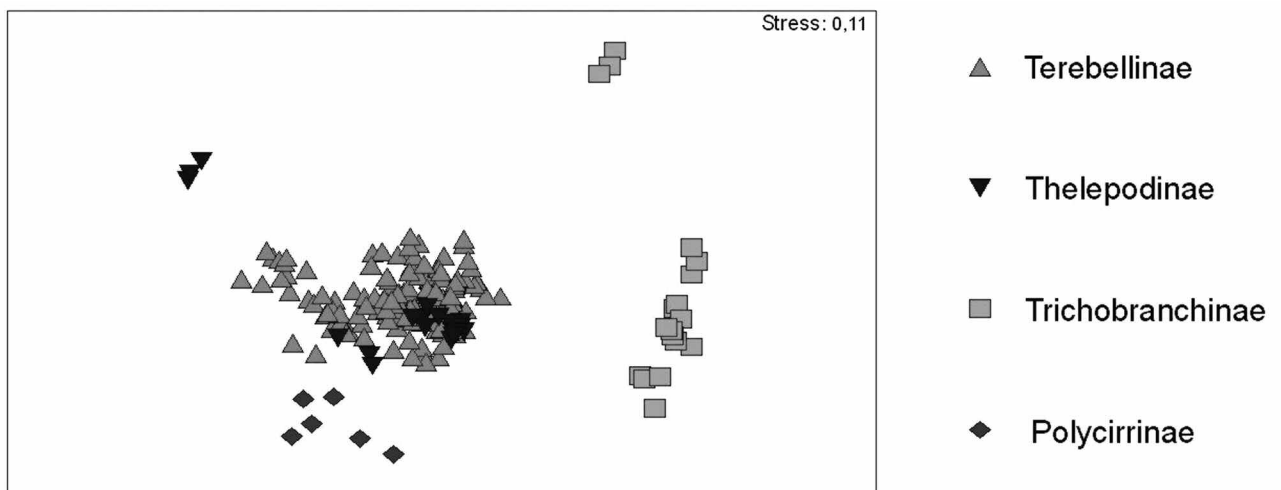
Figure 3. Two-dimensional plot of 17 measurements on three different uncini from segment 16 analysed by non-metric multi-dimensional scaling.
Figure 3. Représentation bidimensionnelle par n-MDS des 17 mesures réalisées sur les uncini du segment 16.

Table 4. Eigenvalues and proportion of the variance explained by the first two principal components of the PCAs realized on the uncini of segments 7 and 16.
Tableau 4. Valeurs propres et part de la variance expliquée par les deux composantes principales des ACP réalisées sur les uncini des segments 7 et 16.

	Eigenvalue	Proportion (%)	Cumulative proportion (%)
Segment 7 (axis 1)	5.890	34.649	-
Segment 7 (axis 2)	5.024	29.555	64.205
Segment 16 (axis 1)	5.573	32.782	-
Segment 16 (axis 2)	5.587	32.277	65.059

In the figure 4 the PCA separation in the uncini from segment 7 was clear between two groups: the Polycirrinae, Thelepodinae and Terebellinae (almost no separation is visible among these groups), and the Trichobranchinae. This distribution was almost equally influenced by the first (34.65%) and second (29.55%) components. In the same way, in the figure 5 the PCA separation in the uncini from segment 16 was very close to what was identified in the Figure 4. This distribution was equally influenced by the first (32.78%) and second (32.27%) components.
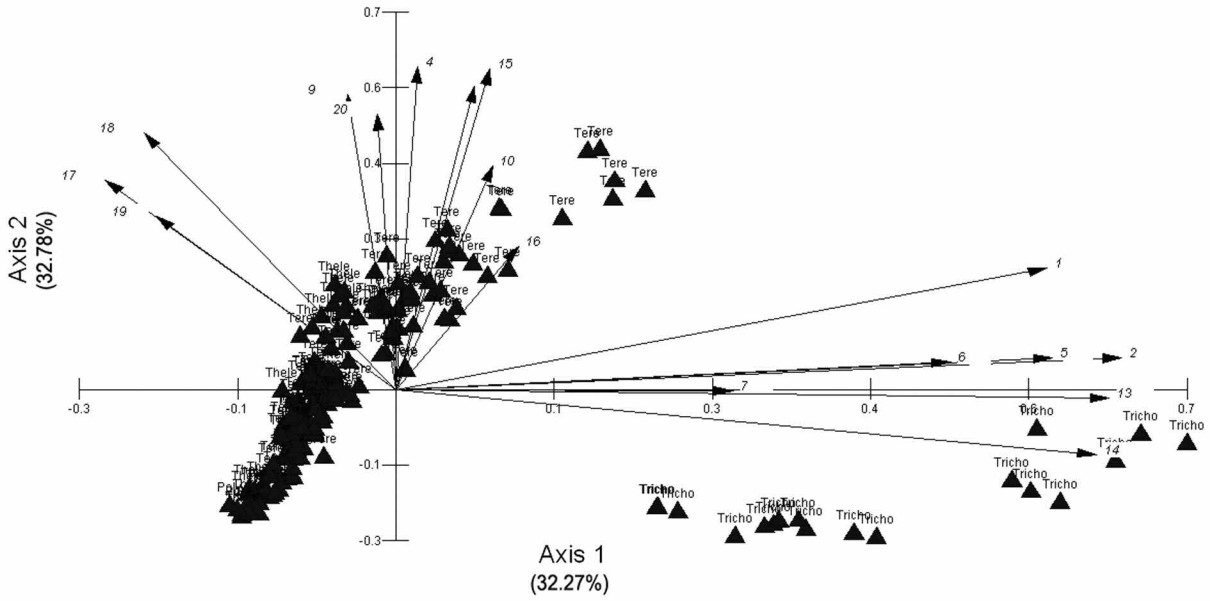
Figure 4. PCA realized on uncini from segment 7 in 31 Terebellidae species.
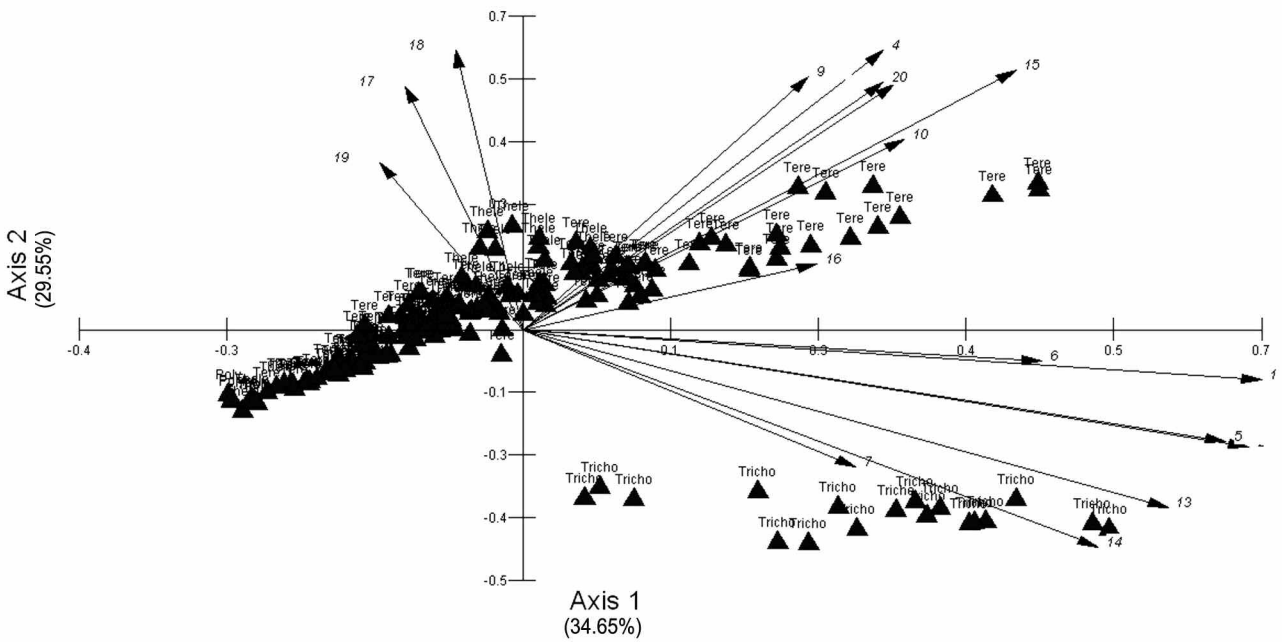**Figure 4.** ACP réalisée sur les uncini du segment 7 chez 31 espèces de Terebellidae.

Figure 5. PCA from segment 16 in 31 Terebellidae species.
Figure 5. ACP réalisée sur les uncini du segment 16 chez 31 espèces de Terebellidae.

Table 5. Coefficient of each variable on the two principal components of the 17 distances measured on segments 7 and 16.

Tableau 5. Coefficient de chaque variable sur les deux composantes principales des 17 distances mesurées sur les segments 7 et 16.

	Axis 1 Seg 7	Axis 2 Seg 7	Axis 1 Seg 16	Axis 2 Seg 16
1	0.385	-0.062	0.373	0.145
2	0.378	-0.144	0.416	0.039
4	0.188	0.343	0.013	0.387
5	0.366	-0.137	0.376	0.038
6	0.27	-0.039	0.318	0.033
7	0.173	-0.169	0.193	-0.002
8	0.193	0.301	0.045	0.365
9	0.149	0.31	-0.028	0.354
10	0.199	0.234	0.055	0.269
13	0.336	-0.218	0.408	-0.01
14	0.3	-0.267	0.401	-0.078
15	0.257	0.319	0.054	0.386
16	0.153	0.081	0.07	0.172
17	-0.061	0.299	-0.166	0.251
18	-0.035	0.343	-0.144	0.308
19	-0.074	0.205	-0.137	0.208
20	0.188	0.304	-0.011	0.33

Table 6. Comparison of the success (%) in the classification at first choice of the four Terebellidae subfamilies by the artificial neural network using uncini from segments 7 and 16 from 31 species.

Tableau 6. Comparaison du succès (%) de la classification à la première assignation des espèces aux quatre sous-familles de Terebellidae par le réseau de neurones, en utilisant les uncini des segments 7 et 16 des 31 espèces étudiées.

		Segment 7	
Number of specimens analysed	**Subfamily**	**Number of the error in the classification**	**Success (%)**
140	Terebellinae	0	100,00
20	Thelepodinae	0	100,00
17	Trichobranchinae	0	100,00
5	Polycirrinae	0	100,00

		Segment 16	
Number of Specimens analysed	**Subfamily**	**Number of the error in the classification**	**Success (%)**
140	Terebellinae	0	100,00
20	Thelepodinae	0	100,00
17	Trichobranchinae	0	100,00
5	Polycirrinae	0	100,00

The PCA showed that almost all distances were important for better data classification (Figs. 4 & 5; Tables 5-8). Thus, unfortunately only a few distances could be removed to improve flexibility in the uncini measurement process. In the PCA of segment 7, only the distances 17, 18, and 19 could be removed from the analysis; on the other hand, on segment 16 only the distances 9, 17, 18, 19, and 20 could be not measured, because their effect on the analysis was weak.

The artificial neural network used here for classification of morphological patterns and prediction of the identification of the subfamily category in Terebellidae, was efficient in detecting the different morphological types, implying that this models can be considered relevant (Table 6). For all the analysed network architectures used on segments 7 and 16 any morphological pattern was incorrect classified, and each subfamily yielded highest percentage in the successful identification. However, the results from the randomization test done with the removed distances due to PCA results, illustrated that the accuracy of the analysis without then was poor and some errors in classification was observed. This misclassication was distributed among all four subfamilies and never was in only one particular subfamily (Table 7).

Discussion

The present study, although it analysed the morphological variation in structures present in individuals grouped together in a high taxonomic category (subfamily), was able to map, group, and classify the different patterns of similarity observed in these structures with success. This result is important to increase systematic knowledge of terebellids, because the primary task of the taxonomist is to understand the morphological variations, either within a taxonomic group or among groups, and to assess the differences among populations or species (Orrhage & Sundberg, 1990).

The success in applying this method was possible because different statistical approaches (linear and non-linear relations) associated with the landmarks were used, which contributed directly to better understanding of the uncini modification in the Terebellidae. The use of statistical methods helped to assess different points related to the delimitation of the linear and nonlinear patterns, and the morphometry led to a better description, grouping the different morphological patterns, and formulating alternatives that facilitated the classification of the new specimens into the morphological patterns observed in categories established prior to the analysis. Finally, the selection of the different chaeta shapes was important to these positive results, because chaetae can be easily dissected, they are chitinous structures, and their shapes are not affected by fixation.

Table 7. Comparison of the success (%) in the classification at first choice of the four Terebellidae subfamilies by the artificial neural network when some distances were excluded. C: correct, E: error.

Tableau 7. Comparaison du succès (%) de la classification à la première assignation des espèces aux 4 sous-familles de Terebellidae par le réseau de neurones quand certaines distances sont exclues. C : correct, E : erreur.

	Segment 7			
Excluded distances	**Terebellinae**	**Trichobranchinae**	**Thelepodinae**	**Polycirrinae**
17	C (100)	C (100)	C (100)	E (00)
17, 18	C (100)	C (100)	C (99)	C (62)
17, 18, 19	C (89)	C (100)	C (99)	C (99)
18	E (18)	C (100)	C (100)	C (96)
18, 19	C (100)	C (99)	E (02)	E (00)
19	C (100)	C (99)	C (96)	C (99)
19, 17	E (00)	C (100)	E (01)	E (00)

	Segment 16			
Excluded distances	**Terebellinae**	**Trichobranchinae**	**Thelepodinae**	**Polycirrinae**
9	E (06)	E (00)	C (54)	E (00)
9, 17	C (100)	E (00)	E (06)	E (00)
9, 17, 18	C(100)	E (00)	C (100)	E (00)
9, 17, 18, 19	E (00)	E (00)	C (100)	E (00)
9, 17, 18, 19, 20	C (100)	E (03)	E (03)	C (96)
9, 17, 19	C (100)	E (00)	E (41)	C (97)
9, 17, 19, 20	C (100)	E (03)	C (99)	C (73)
9, 17, 20	C (67)	C (79)	E (00)	C (99)
9, 18	E (00)	C (99)	C (100)	C (99)
9, 18, 19	E (00)	C (100)	C (100)	C (99)
9, 18, 19, 20	C (64)	C (95)	E (00)	E (36)
9, 18, 20	E (23)	C (98)	E (30)	C (74)
9, 19	C (83)	C (100)	C (55)	C (87)
9, 19, 20	C (100)	C (99)	C (95)	C (100)
9, 20	C (100)	C (100)	C (70)	C (99)

The results presented here are very close to the findings of Garraffoni & Camargo (2006), both by measuring some distances directly from the illustrations of chaetae in the species descriptions, and by dissecting uncini from different species. These results indicate once more that morphometric tools are very efficient to delimit the different boundaries among the different character states, because these tools were able to distinguish groups from variations that at first were considered as overlapping.

Distinct morphological patterns

The present results further strengthened the conclusions by Garraffoni & Camargo (2006) in relation to the identification of three different uncini patterns, and the absence of homology between the long shaft manubrium present in species of Trichobranchinae, and the posterior process present in some species of Terebellinae. It was also possible to establish that the variation in uncini shape can be also coded as three distinct characters (1- character related to the manubrium size: short or long shaft manubrium; 2 - character related to the development of the posterior

process: present or absent; 3- character related to the overall uncini shape, with the three main morphological variations in this chaeta: Trichobranchinae-like, Polycirrinae-like, or Thelepodinae-Terebellinae-like), all of which could be used in a future cladistical analysis of the family. It is important to note that although Glasby & Glasby (2006) reported that they found two types of uncini present in the genus *Polycirrus* (subfamily Polycirrinae), in our analysis, using higher taxonomical level, we found only one pattern that represented the whole subfamily.

Measurements used in the present paper

We note the accuracy of the classification of the group established before the analysis, and the number of the distances measured used in the prediction model. This result is directly related to three main factors: 1) a large number of specimens were measured in each of the four subfamilies, representing a good selection of the morphological patterns; 2) a large number of measurements; 3) a large quantity of replicates for each segment (Morgan et al., 1998). The PCA analysis showed that

almost all components (distances) used in the present study were important to produce good precision of the results. Few distances could be removed, to facilitate data collection without losing precision in the final result. However, when tested, the accuracy of the program after removal of some distances that were indicated by the PCA as not important to the model, negatively affected the prediction model, with loss of precision. Thus, although the PCA assigned less importance to a few distances, it was best to use all of the 17 measurements present in the model, because, probably, careful selection of the distances measured in the uncini was important to the identification of morphological variation in the chaetae.

Artificial Neural Network

Among the different analytical techniques that are currently used to identify a specific specimen in a certain Linnaean category, the artificial neural network is one of those which give the best results (Weeks et al., 1997). This good result occurs because the methodology gives researchers the ability to develop a single, integrative model for predicting the taxon identification from a common set of measurements done in a morphological structure. Neural networks are powerful pattern-recognition and data-analysis tools which make no assumptions about data, offering much more flexibility in statistical modeling (Olden & Jackson, 2002). However, although this is a modern and efficient technique with high potential, this tool has not yet been used in taxonomic studies for species identification (Morgan et al., 1998; Dobigny et al., 2002).

The results of the program based on the neural networks indicated that it is a potentially useful tool to help researchers identify taxonomic groups, because the test carried out here showed good accuracy in prediction of the four subfamilies analysed. However, the final result does not have very strong support, because we used a high hierarchical level (subfamily). On the other hand, we stress that the number of specimens that were included in the data matrix was fundamental to determine the degree of refinement of the taxa that were classified. Depending on the quantity and the diversity of the data included, the actual model would be used to make predictions about hierarchical levels lower than subfamily that will increase the precision and accuracy of terebellid identification for researchers who are not specialists in polychaetes taxonomy.

Conclusions

Morphometry is a powerful tool in the study of morphological variation, mainly at the species level, because it helps to understand the microevolution and the delineation of morphotypes, which will clearly improve the systematic understanding of the Terebellidae.

Some authors have reported that certain terebellid groups (e.g., Terebellinae and Thelepodinae) may be, in fact, non-monophyletic (Brown et al., 1999; Colgan et al., 2001; Garraffoni & Lana, 2004; Glasby et al., 2004). The main problem with the approach used here lies in this possibility, because in this approach the taxonomic categories must be delimited prior to the analysis. Thus, if in a future phylogenetic analysis one of the taxa used in the present paper presents no apomorphic feature, and come to be considered a non-natural group, the model will need to be changed. However, the character states defined by Garraffoni & Camargo (2006) and in the present paper will not be affected if one or more taxonomic categories are altered, because, in reality, the general uncinal shapes and the components used to discriminate them are independent of the Linnaean classification.

References

Ben-Eliahu M.N. 1987. An approach to Nereidid morphometry. *Biological Society of Washington Bulletin*, **7**: 169-173.

Becerra J.M & Valdecasas A.G. 2004. Landmark superimposition for taxonomic identification. *Biological Journal of the Linnaean Society*, **81**: 267–274.

Bookstein F.L. 1991. *Morphometric tools for landmark data.* Cambridge, Cambridge University Press, 435 pp.

Brown S., Rouse G.W., Hutchings P.A. & Colgan D. 1999. Assessing the usefulness of histone H3, U2 snTNA and 28SrDNA in analysis of polychaete relationships. *Australian Journal of Zoology*, **47**: 499-516.

Carvalho A.P.L.F. 2005. Redes neurais artificiais. (http://www.icmc.usp.br/~andre/research/neural/desenv.htm).

Chimimba C.T., Dippenhaar N.J. & Robinson T.J. 1999. Morphometric and morphological delineation of southern African species of *Aethomys* (Rodentia: Muridae). *Biological Journal of the Linnaean Society*, **67**: 510–527.

Chui Y.W., Chen H.C., Lee S.C. & Chen C.A. 2002. Morphometric analysis of shell and operculum variations in the viviparid snail *Chipangopaludina chinensis* (Mollusca: Gastropoda) in Taiwan. *Zoological Studies*, **41:** 321-331.

Clarke K.D. & Warwick R.M. 1994. *Change in marine communities: an approach to statistical analysis and interpretation.* Plymouth: Plymouth Marine Laboratory, 144 pp.

Colgan D., Hutchings P.A. & S. Brown 2001. Phylogenetic rela-

tionships within the Terebellomorpha. *Journal of the Marine Biological Association of the United Kingdom*, **81**: 765-773.

Costa L.F., Reis S.F., Arantes R.A.T., Alves A.C.R & Mutinari G. 2004. Biological shape analysis by digital curvature. *Pattern Recognition*, **37**: 515-524.

Costa-Paiva E.M. & Paiva P.C. 2007. A morphometric analysis of *Euncie cuvier* (Annelida, Polychaeta) species. *Revista Brasileira de Zoologia*, **24**: 353-358.

David B. & Laurin B. 1998. Geometric morphometrics approaches to biodiversity. *Océanis*, **24**: 287-303.

Dedecker A.P., Goethals P.L.M., Gabriels W. & Pauw N. 2004. Optimization of Artificial Neural Network (ANN) model design for prediction of Macroinvertebrates in the Zwalm river basin (Flanders, Belgium). *Ecological Models*, **174**: 161-173.

Dobigny G., Aniskin V. & Volobouev V. 2002. Explosive chromosomal evolution and speciation in the gerbil genus *Taterillus* (Rodentia, Gerbillinae): A case of two new cryptic species. Cytogenet. *Genome Research*, **96**:117-124.

Fauchald K. 1991. A morphometric study of eunicid polychaetes from Belize, Western Caribbean Sea. *Ophelia*, **5**: 47-53.

Garraffoni A.R.S. & Lana P.C. 2004. Cladistic analysis of the subfamily Trichobranchinae (Polychaeta, Terebellidae). *Journal of the Marine Biological Association of the United Kingdom*, **87**: 973-983.

Garraffoni A.R.S. & Camargo M.G. 2006. First application of morphometrics in a study of variations in uncinial shape present within the Terebellidae (Polychaeta). *Zoological Studies*, **45**: 75-80.

Glasby C.J., Hutchings P.A. & Hall K. 2004. Assessment of monophy and taxon affinities within the polychaete clade Terebelliformia (Terebellidae). *Journal of the Marine Biological Association of the United Kingdom*, **84**: 961-971.

Glasby C.J. & Glasby T.M. 2006. Two types of uncini in Polycirrus (Polychaeta: Terebellidae: Polycirrinae) revealed using geometric morphometrics. *Journal of Natural History*, **40**: 237-253.

Guerrero J.A., Luna E. & Sánchez-Hernández C. 2003. Morphometrics in the quantification of character state identity for the assessment of primary homology: an analysis of character variation of the genus *Artibeus* (Chiroptera: Phyllostomidae). *Biological Journal of the Linnaean Society*, **80**: 45-55.

Jordaens K.S., Van Dongen P., Van Riel S., Geenen R., Verhagen T. & Backeljau T. 2002. Multivariate morphometrics of soft body parts in terrestrial slugs: comparison between two datasets, error assessment and taxonomic implications. *Biological Journal of the Linnaean Society*, **75**: 533-542.

Kligenberg C.P. 2002. Morphometrics and the role of the phenotype in studies of the evolution of developmental mechanisms. *Gene*, **287**: 3-10.

MacLeod N. 2001. Landmarks, localization, and the use of morphometrics in phylogenetic analysis. *Fossils, phylogeny, and form: an analytical approach*. (G. Edgecombe, J. Adrain, B. Lieberman eds.), pp. 197-233. New York, Kluwer Academic/Plenum.

Maltagliati F., Camilli L., Lardicci C. & Castelli A. 2001. Evidence for morphological and genetic divergence in *Perinereis cultrifera* (Polychaeta: Nereididae) from two habitat types at Elba Island. *Journal of the Marine Biological Association of the United Kingdom*, **81:** 411-414.

Medeiros L.F. 2003. Redes Neurais em Delphi. Florianópolis, Visual Books Editora, 115pp.

Monteiro L.R. & Reis S.F. 1999. Princípios de morfometria geométrica. Ribeirão Preto, Holos Editora, 188pp.

Monteiro L.R., Diniz-Filho A.F., Reis S.F. & Araújo E.D. 2002. Geometric estimates of hereditability in biological shape. *Evolution*, **56**: 563572.

Morgan A., Boddy L., Mordeu E.M. & Morris C.W. 1998. Evaluation of artificial neural networks for fungal identification, employing morphometric data from spores of *Pestalotiopsis* species. *Mycology Research*, **102**: 975-984.

Olden J.D. 2000. An artificial network approach for studing phytoplankton succession. *Hydrobiologia*, **436**: 131-143.

Olden, J.D. & Jackson D.A. 2002. A comparison of statistical approaches for modelling fish species distributions. *Freshwater Biology*, **47**: 1976-1995.

Orrhage L. & Sundberg P. 1990. Multivariate analysis of morphometric differentiation within the *Laonice cirrata*-group (Polychaeta: Spionidae). *Zoologica Scripta*, **19**: 173-178.

Reis S.F. 1988. Morfometria e estatística multivariada em biologia evolutiva. *Revista Brasileira de Zoologia*, **5**: 571-580.

Rolf F.J. & Marcus L.F. 1993. A revolution in morphometrics. *Trends in Ecology and Evolution*, **8**: 129-132.

Roth V.L. & Mercer J.M. 2000. Morphometrics in development and evolution. *American Zoologist*, **40**: 801-810.

Rumelhart D., Hinton G. & Williams R. 1986. Learning internal representations by back-propagating errors. *Nature*, **323**: 533-536.

Tatibana C.Y. & Kaetsu D.Y. 2005. Una introdução a rede neural (http://www.din.uem.br/ia/neurais).

Weeks P.J.D., Gauld I.D., Gaston K.J. & O'Neill M.A. 1997. Automating the identification of insects: a new solution to an old problem. *Bulletin of Entomological Research*, **87**: 203–211.