

PREPRINT

© 1998 Koninklijke Brill, Uitgevers.

A procedure to extract phylogenetic information from morphometric data

Jeroen van Velzen, Niels Bouton and Rino Zandee
Institute of Evolutionary and Ecological Sciences, University of Leiden,
PO Box 9516, 2300 RA, Leiden, The Netherlands.

Corresponding author:

Niels Bouton

address: Institute of Evolutionary and Ecological Sciences, PO Box 9516, 2300
RA Leiden, The Netherlands

Fax: +31 71 5274900

E-mail: Bouton@rulsfb.leidenuniv.nl

date: 1998-07-06

To reconstruct a phylogeny, data with sufficient resolving power are necessary. When morphological or molecular data fail this criterion, other data types are needed.

Morphometric characters can sometimes offer a solution. However, environmental effects due to differences in location have to be eliminated or at least reduced before these characters can be applied in phylogeny reconstruction. Here we present a new procedure to extract phylogenetic information from morphometric characters. The procedure is based on assigning weights to these characters.

The procedure starts with a principal component analysis of the morphometric measurements. Then an analysis of variance (ANOVA) is done on principal components scores in order to determine species effects, environmental effects due to differences in location, and interaction effects. Subsequently, components are chosen for weighting characters on the grounds of two criteria. First, there must be only a significant species effect on their scores. Second, they have to account for a large proportion of total variance. The measurements are discretized by the homogeneous subset coding procedure. The discretized characters are weighted by means of the proportions of variance that are contributed to the chosen PC's.

We used head measurements of haplochromines to test the procedure. Evaluations indicated that our procedure is capable of extracting phylogenetic information from morphometric data.

keywords: morphometric data, phylogeny, PCA, haplochromines, cichlids

Introduction

The use of morphometric data to reconstruct phylogenetic relationships is complicated by environmental effects due to differences in location on the measurements.

Phylogenetic and non-phylogenetic information have to be separated to be able to concentrate solely on phylogenetic information. We developed a new procedure to extract phylogenetic information from morphometric measurements for phylogenetic reconstructions. This new procedure is a combination of already available analytical methods, namely principal component analysis, analysis of variance and the homogeneous subset coding procedure. Next to extraction of phylogenetic information some other aspects concerning the use of morphometric data are addressed here.

We tested the procedure by applying it to a group of haplochromine cichlids. The haplochromine species flock is presumably monophyletic (Meyer *et al.*, 1990; Meyer, 1993). Both morphological and molecular characters lacked resolution to reconstruct phylogenies on species level. However, morphometric data are available.

Bouton *et al.* (manuscript) investigated the anatomy of the head region as reflected in the outer head shape. Local selection pressures lead to changes in the anatomy of the apparatuses in the head, resulting in variation in the outer head shape between populations of the same species. They conclude that differences in the head are the result of phenotypic plasticity, local adaptation and phylogenetic relationships.

Different evaluations of this example showed that the procedure is able to extract phylogenetic information from morphometric data and minimize the influences of local selection pressures on the results of phylogeny reconstruction.

Outline of the procedure

The procedure outlined below aims at extracting phylogenetic information from morphometric data. It consists of five main steps:

- 1 - elimination of growth (allometric) trends from the data (standardization),

- 2 - elimination of environmental effects (PCA + ANOVA),
- 3 - detecting species delimitation (ANOVA),
- 4 - deriving character weights (PCA),
- 5 - discretizing morphometric characters (homogeneous subset coding).

Character standardization

Besides effects from the environment and evolutionary history, morphometric characters may contain growth and/or allometric trends. To correct for (relative) differences in size all measurements should be standardized, e.g. expressed as proportions of a chosen standard length. Moreover, selecting only specimens from a specific size range may also contribute in the elimination of growth trends.

Principal component analysis (PCA)

A principal component analysis is performed on morphometric characters (Morrison, 1976; Pimentel, 1979; Bookstein *et al.*, 1985) using a covariance matrix.

Analysis of variance (ANOVA)

Three effects are expected to influence the morphometric data: a species effect, a location effect, and an interaction between species and locations. A standard analysis of variance is done on the scores for each principal component (PC) carrying more than 5% of the total variation to explore the influences of these effects. We are interested in PC's that show a significant species effect and no significant effect for location and species-location interaction. Of these PC's we use the one explaining most of the original variation.

Deriving character weights

Each character contributes to the different PC's according to its loading for that PC. The total variance of a character is consequently divided over all PC's. The proportion of the variance which is contributed to the PC that is of our interest is used as a weight in the phylogenetic analysis. These proportions are a direct reflection of the relative contribution of a character to an important aspect of species differentiation.

Coding procedure

It is necessary for phylogeny reconstruction to have discrete characters. The morphometric data are discretized according to the homogeneous subset coding procedure of Simon (1983). Homogeneous subsets were detected through the application of the multiple range test (Miller, 1966: p. 81; significance level = 0.05). We considered the character states unpolarized and linearly ordered *a priori* on the assumption that transition can only take place between neighbouring character states. Not all characters have to be coded within the same range. Characters that show significant differences between more species have a larger range in their coding. These characters receive a higher number of discrete states and consequently have a larger impact on the phylogeny reconstruction.

Phylogeny reconstruction

The data matrices obtained by the homogeneous subset coding procedure and the weighting of the characters determined by the principal component analysis are used to reconstruct the most parsimonious network (MPN) according to the parsimony criterion (Kluge & Farris, 1969). There are several algorithms to find this MPN. We use PAUP version 3.0 (Swofford, 1990).

Methods and Materials

The procedure described above was used to reconstruct the phylogeny of rock-dwelling haplochromine cichlids. Fishes were sampled at four rocky islands in the southern part of Lake Victoria (figure 1): Python and Anchor Islands in the Mwanza Gulf, and Makobe and Ruti Islands in the Speke Gulf.

[insert figure 1]

The following six species were included: *Pundamilia nyererei* (Witte-Maas & Witte, 1985), *Neochromis rufocaudalis* (Seehausen & Bouton, 1998), *Neochromis greenwoodi* (Seehausen & Bouton, 1998), *Pundamilia macrocephalus* (Seehausen & Bouton, 1998), *Mbipia mbipi* (Seehausen, Lippitsch & Bouton, 1998), and *Neochromis omnicaeruleus* (Seehausen & Bouton, 1998). In total, twelve populations were used (table I).

[Insert table I.]

The external framework of the head (De Visser & Barel, manuscript) was measured on five specimens of each population. The external framework consists of thirteen points on the head surface (appendix 1, figure 2), describing the outer head shape in which an oral, a suspensorial and an opercular compartment are distinguished. De Visser & Barel (manuscript) state that the measurements are homologous.

[insert figure 2]

Nineteen measurements (appendix 2) were taken on a continuous scale to the nearest 0.1 mm using digital callipers. On the assumption of bilateral symmetry, paired distances were measured only at the left side. To minimize effects of allometry, five adult males of similar size of each population were selected. The standard length ranged from

between 8.7 and 9.1 cm. Measurements were standardized by expressing them as proportions of the length axis of the medial plane of the neurocranium (De Visser & Barel, manuscript).

Results

Principal component analysis

The first five PC's accounted for respectively 32.8, 24.6, 7.3, 7.2 and 5.2 percent of the total variance (table II). The variance load of the first PC could be considered low for this kind of data. The first PC usually accounts for more than 80 percent of the total variance (*e.g.* Reinthal, 1990). In our study the specimens were all of similar size and the measurements were standardized. Therefore, no growth trend was expected on the first principal axis. Individual scores on the first and second PC (figure 3) indicated a differentiation mainly among locations along the first principal axis, and mainly among species along the second principal axis.

[insert figure 3]

Analysis of variance

We estimated the effects of species, location and the interaction of species and location on the principal component scores. In this analysis of variance we disregarded the PC's that described less than 5% of the total variance (table II). The second as well as the fifth PC showed a significant influence of the species effect and no significant influences of location effect or interaction of species and location. The second PC described a larger part of the total variation and was used to obtain character weights for phylogenetic analysis.

Deriving character weights

The weights were determined by the ratio between the loading on the second PC and the sum of the loadings for all PC's.[Insert table II]

The second PC was mainly determined by the feeding apparatus. The three characters that had the largest proportion of their variances explained by the second PC (table III; LJS-QHC, SLF-SLF, QHC-QHC) are all localised in the feeding apparatus (Bouton *et al.*, manuscript). This reflects an adaptation to the feeding behaviour of the species.

Coding procedure

Two homogeneous subset coding procedures were performed: one to obtain scores at species level, and one to obtain scores at population level. At species level, the ranges of the scores varied from two to six (table IV). At population level, the ranges varied from three to twelve (table V).

[Insert table IV and table V]

Phylogenetic analysis

The character weights for phylogenetic analysis were determined by the proportion of variance that the character contributed to the second PC. Table III shows the weighting factors and the loadings on the second PC for each character.

The standardized, discretized and weighted measurements were used to perform a phylogenetic analysis of the six species. Figure 4 shows the MPN that was found.

[insert figure 4]

Evaluation

The procedure was evaluated in three ways. First, an MPN was reconstructed using populations as taxonomical units. Second, we performed a phylogenetic analysis at population level in which all character changes were given equal weights. Third, we mapped other data on our MPN of the species.

We performed a phylogenetic analysis at population level using the scores that were obtained by the homogeneous subset coding procedure at population level. The weights were already determined by our procedure on species level. This weighting scheme should minimize differences between populations of the same species, and consequently should group populations of the same species as a monophyletic group. The analysis resulted in one MPN (figure 5). Populations of three species were grouped together, indicating that the procedure successfully corrects for influences of locations and species-location interactions. One species was represented by one population only and can therefore not be used for this evaluation. The *M. mbipi* and *N. omnicarulaeus* populations were not grouped together. This indicated that, on average, the variation between these populations (and within species) is higher than the variation among the corresponding species.

[insert figure 5]

We also performed a phylogenetic analysis at population level with equal weights for all characters to examine the effects of our procedure. The MPN was expected to show an overestimation of phylogenetic relatedness between populations of one location. The analysis resulted in four equally parsimonious networks of which we show the strict consensus (figure 6).

[insert figure 6]

When this MPN was rooted between *M. mbipi* of Ruti island and *N. rufocaudalis* of Python Island, the cladogram split up populations according to the gulf where they were caught (Python Island and Anchor Island in the Mwanza Gulf and Makobe Island and Ruti Island in the Speke Gulf). The differences between populations of the same species are possibly a reflection of the differences in the physical environment of the gulfs (Bouton *et al.*, manuscript). The only exception is *P. nyererei* for which populations from Python Island and from Ruti Island are indicated as one group. Obviously these populations of *P. nyererei* differ less from each other than they differ from the populations of other species from the same location.

Finally, we mapped other data on the MPN resulting from our procedure. Three colour characters (colour of flank and dorsum, colour of dorsal fin and colour of anal fin) and two teeth characters (shape of outer row teeth and number of inner rows) were available for this purpose. The number of character state changes necessary to fit the colour and teeth characters on the MPN based on the head measurements was 26, whereas the most parsimonious solution was 24 steps. When we compared the number of character state changes among all possible networks, our network was within the set of the 10% most parsimonious solutions.

Discussion

Morphometric data can be affected by both evolutionary history and recent environmental influences. These effects have to be distinguished for phylogeny reconstruction. In this paper we propose a new procedure to achieve this.

Bookstein (1994) argues that morphometric characters are not suited at all to be used in phylogeny reconstruction because they are not independent. They have a functional relationship and describe a non-euclidean geometry. This may cause overestimation of the phylogenetic information content of the data. However, if these kinds of dependencies are associated with species delimitation, they contain an evolutionary historical component, which can be detected with our procedure and subsequently be used in a phylogeny reconstruction.

Coding procedure

Pimentel and Riggins (1987) argue that quantitative characters are not suitable for phylogenetic analysis. They see no way to interpret discretized quantitative characters. They assume characters to change non-gradually. Therefore they can apply strong criteria to determine character states. However, this assumption is not necessary to determine character states for phylogeny reconstruction. In this case a weaker statistical criterion is sufficient to determine character states. Here we showed that discretized quantitative characters are suitable for phylogenetic analysis and permit a sensible biological interpretation in terms of trophic types. In this we agree with Thorpe (1984). Converting characters from a continuous to a discrete scale causes, in general, loss of information. In our case, this is no problem since a high number of significant differences between taxonomical units were preserved.

Archie (1985) describes different methods to project continuous characters on a discrete scale. The homogeneous subset procedure is considered to have the disadvantage that the range on which character states are scored, is variable. The range

determines the relative importance of a character in the phylogenetic analysis. The number of steps needed to fit ordered characters with a large range onto a cladogram will vary more than characters with a smaller range. Therefore, the characters with a relatively large range will have a larger impact on the choice of the shortest cladogram. In this study only morphometric characters were used, all discretized by the homogeneous subset procedure. Differences in coding range reflect the number of significant differences between species. Characters with a larger range described differences between species easier and therefore are assumed to contain a larger amount of phylogenetic information.

Alternative methods, like the gap-coding procedure (Mickevich & Johnson, 1976) determine discrete states by searching for gaps on the continuous scale. Any pair of adjacent means is given a different score if these means are separated by more than the within-group standard deviation times an arbitrary constant. The generalized gap-coding procedure (Archie, 1985) determines groups of means, under the constraint that no pair of members of one group may differ more than the within-group standard deviation times an arbitrary constant. Both gap-coding procedures assume the variance of the estimated means to be equal for all species. Here this cannot be assumed, for two reasons. Firstly, the number of specimens per species was not constant. Secondly, the variance is expected to be larger when the mean is larger. As the assumptions of the gap-coding procedures could not be met, we used the homogeneous subset procedure.

Determination of weights

The second PC accounted for a considerable percentage of the total variation (24.6 %) and ANOVA showed that it was highly determined by differences between species. Therefore, this PC was used to determine the weight of the characters. An alternative would be to include the fifth PC, which also was determined by species effects. However, we decided not to do this because the fifth PC explains only just

above 5% of the total variance and moreover its locality effect is much larger (16%) than that of the second PC.

The procedure can be generalised when more than one PC carries a significant species effect. In that case the weight per character can be summed over PC's.

We also considered minimizing location effects by choosing weights inverse to the proportion of variance explained by PC's that are determined by location effects.

However, in our case none of the PC's that accounted for more than 5 percent of the total variation showed a pure location effect (without species/location interaction; see table II).

Evaluation results

The phylogenetic analysis at population level based on weighted characters confirmed that our procedure is able to extract phylogenetic information from the data. From this evaluation no conclusion about the conspecificity of the populations can be drawn because the procedure was designed to minimize differences within populations of the same species.

The phylogenetic analysis at population level based on unweighted characters showed that most variation in the data set is caused by plasticity in reaction to habitat differences between the gulfs. The first PC showed a significant location effect (table II) and described mainly differences between populations of different gulfs (figure 3a).

This clearly shows the need to focus on interspecific differences as done with our procedure.

To fit colour and teeth characters on the MPN for head measurements 26 character state changes were needed, whereas 24 was the most parsimonious solution. Because of the low number of species, we were able to determine the exact distribution of tree lengths for all possible trees. The fitting by 26 steps was within the best 10%. In general, a 5% upper limit is used. However, in bootstrap analysis (Felsenstein, 1985) it is often found that subsets of characters do not carry enough information to reconstruct the phylogeny that is obtained when the total data set is used. Reconstructions obtained from small

data sets, such as the set of the colour and teeth data we used here, tend to have a larger uncertainty.

In conclusion, the evaluations indicate that this procedure enables us to extract phylogenetic information from morphometric data, provided that samples of sufficient size from different locations are available. An alternative procedure could consist of a multivariate analysis of variance MANOVA with two factors, species and locations. However, to be able to perform this MANOVA a considerable sample size per population is necessary relative to the number of characters studied. In this case only five specimens were available, so we could not test the effectivity of this alternative. The small number of specimens per cell also restrained us to use pooled within-group covariance matrices. As we demonstrated, our procedure as proposed has the advantage over these alternatives that a smaller sample size per location suffices.

A definite confirmation of the phylogeny of the rock-dwelling haplochromine cichlids has to wait till an extensive set of characters, such as DNA sequences, becomes available.

Acknowledgements

We thank Patsy Haccou for advice and comments on previous versions of the manuscript. Furthermore we thank J. de Visser, J.J.M. van Alphen, J.G. Sevenster and O. Seehausen for stimulating discussions and M. Brittijn for drawing the figures. This research was supported by a WOTRO grant W84-282 to N. Bouton.

References

- Archie, J.W., 1985. Methods for coding variable morphological features for numerical taxonomic analysis. *Syst. Zool.*, **34**: 326-345.
- Bookstein, F. L., B. Chernoff, R. L. Elder, J. M. Jr. Humphries, G. R. Smith, & R. E. Strauss, 1985. Morphometrics. In: *Evolutionary Biology*, Special Publication 15, The Academy of Natural Sciences of Philadelphia, Philadelphia.
- Bookstein, F.L., 1994. Can biometrical shape be a homologous character? In: B.K. Hall (ed.) *Homology, The hierarchical basis of comparative biology*, Academic Press, Halifax.
- Bouton, N., C.D.N. Barel, J. de Visser, & O. Seehausen, manuscript. Differences in head shape among populations of rock-dwelling haplochromine species from Lake Victoria with a discussion on ecological significance.
- De Visser, J. & C.D.N. Barel, manuscript. Functional interpretation of a 3-D truss analysis of cichlid head shapes.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using bootstrap. *Evolution*, **39**(4): 783-791.
- Kluge, A.G. & J.S. Farris, 1969. Quantitative phyletics and the evolution of anurans. *Syst. Zool.*, **18**: 1-32.
- Meyer, A., 1993. Phylogenetic Relationships and Evolutionary Processes. In: *East African Cichlid Fishes*. TREE **8**(8): 279-284.
- Meyer, A., T.D. Kocher, P. Basasibwaki, & A.C. Wilson, 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* **347**: 550-553.
- Mickevich, M.F. & M.F. Johnson, 1976. Congruence between morphological and allozyme data in evolutionary inference and character evolution. *Syst. Zool.*, **25**: 260-270.
- Miller, R.G. jr., 1966. *Simultaneous statistical inference*. Mcgraw-Hill, New York.
- Morrison, D.F., 1976. *Multivariate Statistical Methods, 2nd ed.* Mcgraw-Hill, New York.
- Pimentel, R.A., 1979. *Morphometrics, The multivariate analysis of biological data*. Kendall/Hunt Publishing company, Iowa.

Pimentel, R.A. & R. Riggins, 1987. The nature of cladistic data. *Cladistics* **3**(3): 201-209.

Reinthal, P.N., 1990. Morphological analyses of the neurocranium of a group of rock-dwelling cichlid fishes (Cichlidae: Perciformes) from Lake Malawi, Africa. *Zool. J. Linn. Soc.* **98**: 123-139.

Simon, C.M., 1983. A new coding for morphometric data with an example from periodical cicada wing veins. In: J. Felsenstein (ed.), *Numerical taxonomy Proceedings of the NATO Advanced Study Institute*: 378-383. NATO Adv. Study Institute Ser. G (ecological Sciences), No. 1. Springer-Verlag, New York.

Seehausen, O., E. Lippitsch, N. Bouton & H. Zwennes, 1998, Mbipi, the rock-dwelling cichlids of Lake Victoria: description of three new genera and 15 new species (Teleostei). *Ichthyological Exploration of Freshwaters*, in press.

Swofford, D.L., 1990. *PAUP, Phylogenetic Analysis Using Parsimony, version 3.0*. Illinois Natural History Survey, Illinois.

Thorpe, R.S., 1984. Coding morphometric characters for constructing distance Wagner networks. *Evolution*, **38**(2): 244-255.

Witte-Maas, E.L.M. & F. Witte, 1985. *Haplochromis nyererei*, a new cichlid fish of Lake Victoria, named in honour of *Mwalimu* Julius Nyerere, president of Tanzania. E. J. Brill, Leiden, 13pp.

Appendix 1 Description of enhanced nomenclature.

Anatomical nomenclature and orientation are according to Barel *et al.* 1977. The suspensorial section of the parasphenoid bar was taken as the horizontal direction.

LJS: The rostral-dorsal tip of the mandibular symphysis; lips to be suppressed.

POP: The ventrolateral corner of the preorbital process, after removing skin and lachrymal.

NPC: The centre of distance between the preopercular top (at the first suspensorial lateral line foramen) and the neurocranial lateral line crest at the fifth neurocranial lateral foramen.

DOP: The caudal dorsal corner of the operculum.

QHC: The centre of the lateral head of the suspensorial mandibular articulation facet (quadrate head).

SLF: The centre of fourth suspensorial lateral foramen.

FDO: The caudal rim of the operculum at the muscular basis of the pectoral fin.

- 1 **LJS-POP**: dorsal oral length
- 2 **LJS-QHC**: ventral oral length
- 3 **POP-NPC**: dorsal suspensorial length
- 4 **POP-QHC**: oral depth
- 5 **POP-SLF**: diagonal 1
- 6 **NPC-QHC**: diagonal 2
- 7 **NPC-SLF**: suspensorial depth
- 8 **NPC-FDO**: opercular depth
- 9 **NPC-DOP**: dorsal opercular length
- 10 **QHC-SLF**: ventral suspensorial length
- 11 **SLF-DOP**: diagonal 3
- 12 **SLF-FDO**: diagonal 4
- 13 **DOP-FDO**: opercular depth
- 14 **POP-POP**: dorsal oral width
- 15 **NPC-NPC**: dorsal suspensorial width
- 16 **DOP-DOP**: dorsal opercular width
- 17 **FDO-FDO**: ventral opercular width
- 18 **SLF-SLF**: ventral suspensorial width
- 19 **QHC-QHC**: ventral oral width

location	Species					
	nye	dee	vel	Cop	nig	blu
Python Island	★	★	★	.	★	
Anchor Island	.	.	★		★	
Makobe Island	.			★	★	★
Ruti Island	★			★		★

Table I. Sample sites and species occurrence. nye = *P. nyererei*, dee = *P. macrocephalus*, vel = *N. greenwoodi*, cop = *M. mbipi*, nig = *N. rufocaudalis*, blu = *N. omnicarulaeus*. ★ sampled in high numbers and used for measurements, · sampled in low numbers.

Principal component	percentage of total variance explained	Interaction Species x location	Species	location
1	32.8 %	0 %	46.5 % **	40.8 % *
2	24.6 %	5.0 %	84.0 % ***	4.6 %
3	7.3 %	21.4 %	27.9 %	26.4 %
4	7.2 %	3.7 %	48.1 %	25.0 %
5	5.2 %	2.2 %	63.3 % **	16.0 %

Table II. Relative amount of variation accounted by the first five principal components and the effects determining their variance. Significance levels: *: p 0.05; **: p 0.01;***: p 0.0005

character	weighting factor	loading second PC
1	0,03	0,694
2	0,77	6,311
3	0,08	-0,540
4	0,13	1,211
5	0,31	1,690
6	0,08	0,855
7	0,01	-0,361
8	0	0,042
9	0,01	-0,276
10	0,17	1,651
11	0,04	-0,857
12	0,16	-1,701
13	0,04	0,872
14	0,05	-0,466
15	0,14	-1,408
16	0	0,021
17	0	0,341
18	0,52	-4,365
19	0,34	-3,457

Table III. Proportions of variance of the individual characters as carried by the second PC. These were used as weighting factors. The third column presents the loadings of the characters on the second PC. For characters descriptions see appendix 2.

species	nye	dee	vel	cop	Nig	blu
character						
1	4	7	5	2	4	1
2	7	7	5	3	1	1
3	3	2	1	2	3	5
4	2	3	2	2	2	1
5	5	5	3	2	2	1
6	4	4	5	1	2	1
7	1	2	3	2	2	1
8	3	5	4	2	4	1
9	1	3	1	1	1	1
10	3	5	5	3	2	1
11	1	5	5	2	4	1
12	1	3	3	3	3	2
13	5	6	7	2	3	1
14	1	5	5	4	4	2
15	1	3	3	1	3	3
16	2	5	3	1	2	1
17	1	3	3	1	1	1
18	1	3	6	5	7	4
19	1	2	4	1	5	1

Table IV. Discrete representation of the head measurements of the species. nye = *P. nyererei*, dee = *P. macrocephalus*, vel = *N. greenwoodi*, cop = *M. mbipi*, nig = *N. rufocaudalis*, blu = *N. omnicaerulaeus*. For characters descriptions see appendix 2.

species	nye	nye	dee	vel	vel	cop	Co p	nig	nig	Nig	blu	blu
location	pt	ru	pt	pt	an	ma	Ru	pt	an	Ma	ma	ru
characters												
1	10	5	13	11	9	1	8	9	6	5	3	2
2	11	11	11	9	8	5	6	2	3	1	3	2
3	3	6	3	1	2	6	2	4	7	8	10	11
4	4	4	5	4	3	4	4	2	4	4	4	1
5	6	6	7	4	4	3	2	4	3	2	3	1
6	5	9	8	8	7	4	2	4	7	3	3	1
7	2	2	4	5	2	3	3	2	3	2	3	1
8	6	2	6	6	4	1	4	5	7	3	3	1
9	2	1	5	2	3	3	2	4	3	2	2	2
10	4	5	9	9	8	4	7	3	5	2	2	1
11	5	2	12	13	10	4	8	9	11	6	3	1
12	2	1	4	5	3	3	4	4	5	3	3	3
13	6	4	8	5	9	2	4	5	6	3	4	1
14	1	2	7	6	6	4	5	6	6	2	5	2
15	2	1	7	5	5	3	1	7	6	5	7	5
16	4	1	9	5	8	3	3	3	3	2	1	3
17	4	1	9	6	8	3	3	3	3	2	1	3
18	2	1	4	7	9	7	7	9	8	7	7	5
19	7	1	5	9	9	2	4	11	9	10	7	4

Table V. Discrete representation of the head measurements of the populations. nye = *P. nyererei*, dee = *P. macrocephalus*, vel = *N. greenwoodi*, cop = *M. mbipi*, nig = *N. rufocaudalis*, blu = *N. omnicarulaeus*. pt = Python Island, an = Anchor Island, ma = Makobe Island, ru = Ruti Island. For characters descriptions see appendix 2.

FIGURE LEGENDS

Figure 1. Map of Lake Victoria and the research area.

Figure 2. Schematic representation of the external framework.

Figure 3a. Individual PC-scores on the first and second principal component, with labels according to location. ■ = Python Island, ● = Anchor Island, □ = Makobe Island, ○ = Ruti Island.

Figure 3b. Individual PC-scores on the first and second principal components, with labels according species. _ = *P. nyererei*, _ = *P. macrocephalus*, ★ = *N. greenwoodi*, ◆ = *M. mbipi*, = *N. rufocaudalis*, = *N. omnicaerulaeus*.

Figure 4. MPN based on weighted head measurements.

Figure 5. MPN at population level, based on weighted head measurements. _ = *P. nyererei*, _ = *P. macrocephalus*, ★ = *N. greenwoodi*, ◆ = *M. mbipi*, = *N. rufocaudalis*, = *N. omnicaerulaeus*. pt = Python Island, an = Anchor Island, ma = Makobe Island, ru = Ruti Island.

Figure 6. The consensus of the MPN's at population level based on unweighted head measurements. _ = *P. nyererei*, _ = *P. macrocephalus*, ★ = *N. greenwoodi*, ◆ = *M. mbipi*, = *N. rufocaudalis*, = *N. omnicarulaeus*. pt = Python Island, an = Anchor Island, ma = Makobe Island, ru = Ruti Island.

Figure 1

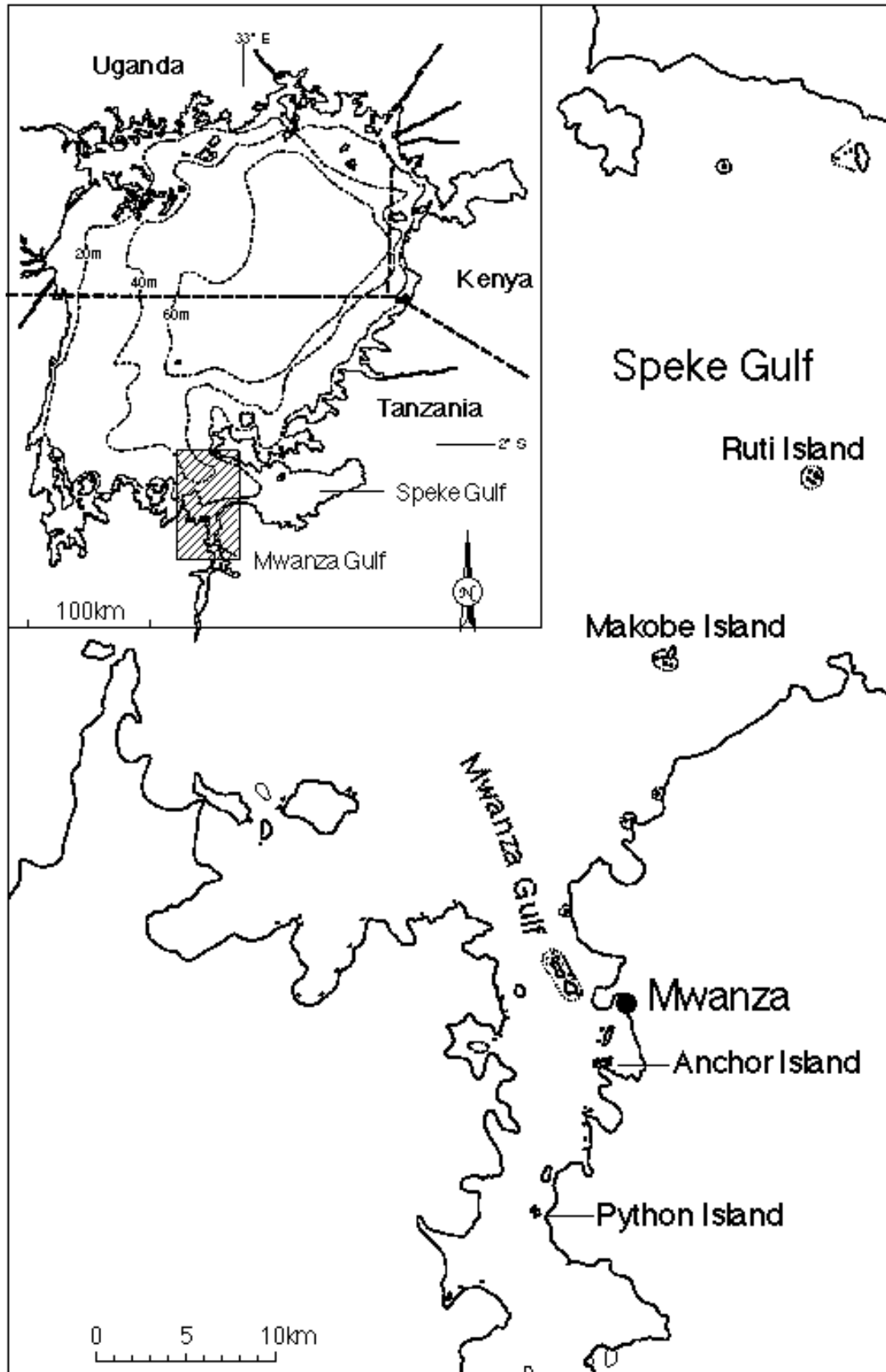
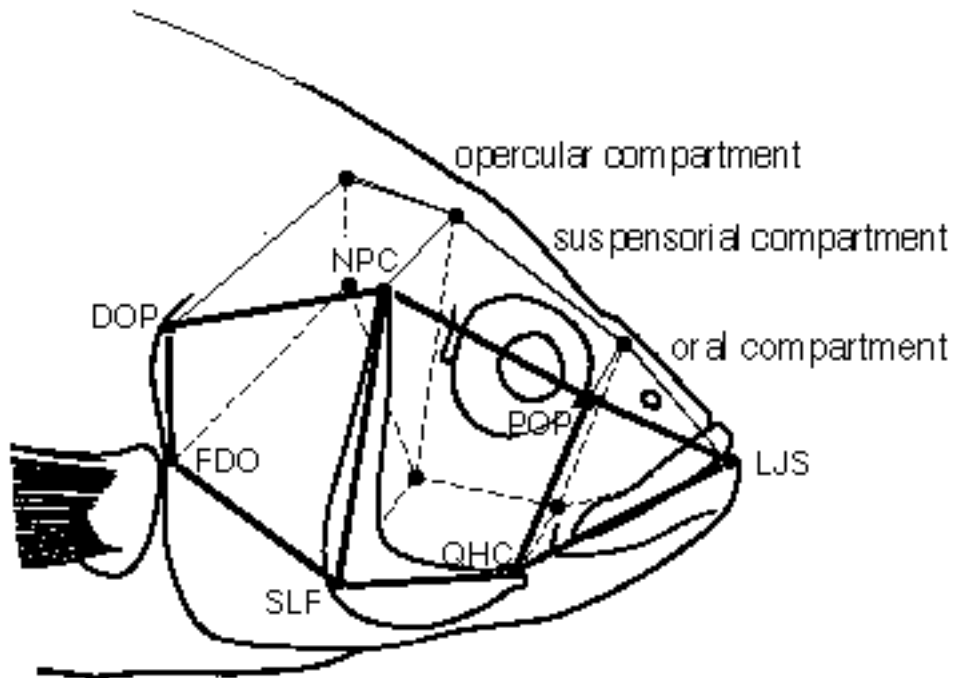


Figure 2



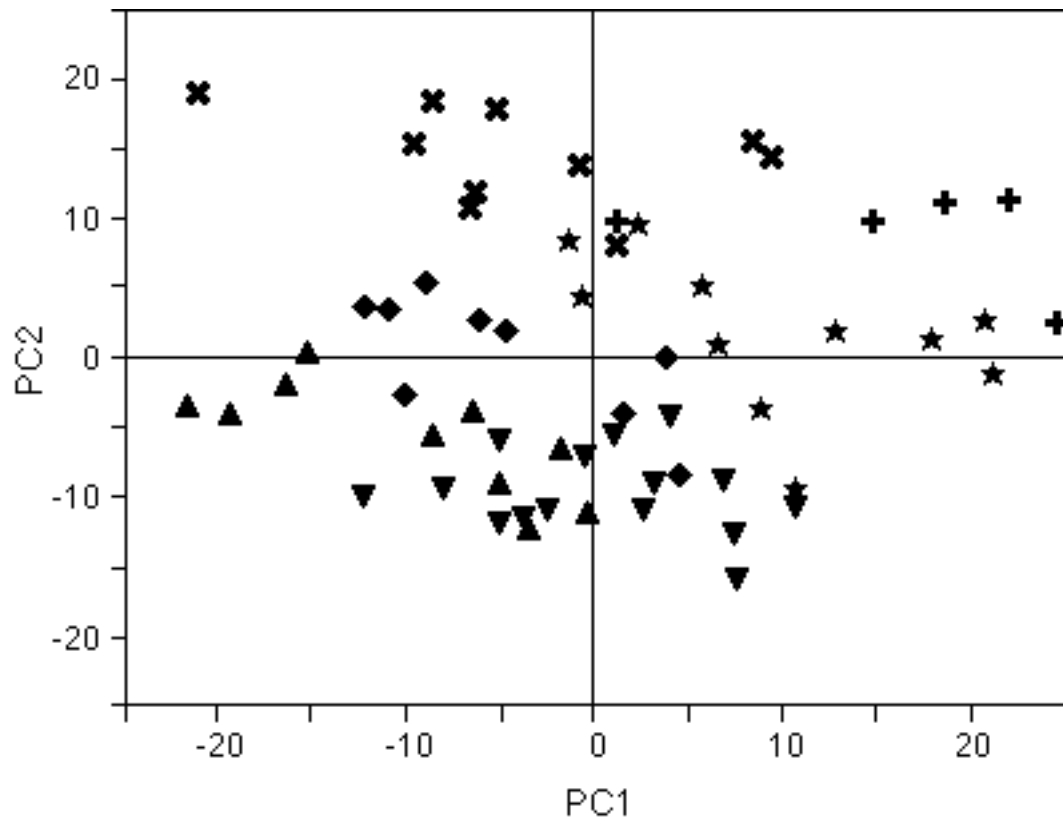


Figure 4

Figure 5

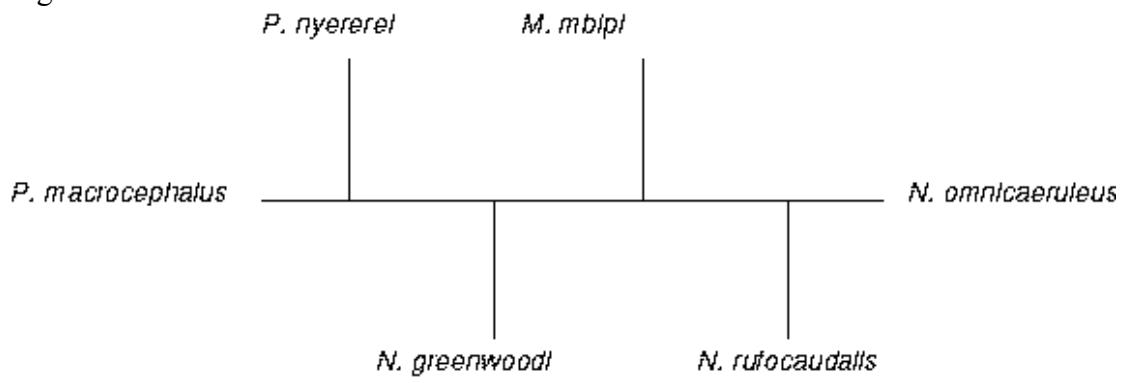


Figure 6

