The Evolutionary Biology and Taxonomic Review of Zambian Riverine Haplochromine Cichlids

By

Cyprian Katongo

BSc (University of Zambia) MSc (University of Birmingham)

A thesis submitted to the University of Zambia in fulfilment of the requirements of the degree of Doctor of Philosophy in Biological Sciences of the University of Zambia.

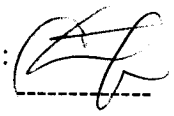
THE UNIVERSITY OF ZAMBIA

LUSAKA

2005

Declaration

I, the undersigned, hereby declare that the work presented in this thesis represents my own work and that it has not been submitted and not being currently submitted in part or whole for the award of a degree at any other university. All sources of material and help have been duly acknowledged.

Signed:  Date: 19-08-2005

Cyprian Katongo

CERTIFICATE OF APPROVAL

THIS THESIS PRESENTED BY CYPRIAN KATONGO ..... **IS APPROVED AS FULFILLING THE REQUIREMENT FOR THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY (BIOLOGICAL SCIENCES) BY THE UNIVERSITY OF ZAMBIA.**

NAME	SIGNATURE
Prof. K.J. Mbata	..... 22/08/2005
Chairperson	
DR L.E. MUMBA	.....
Supervisor and Internal Examiner	
DR. P.M. CHIMUNGU	.....
Internal Examiner	
PROF. A.J.D. AMBALI	.....
External Examiner	
N/A	N/A
.....	.....
Member	
N/A	N/A
.....	.....
Member	

22/08/2005

Abstract

The Evolutionary Biology and Taxonomic Review of Zambian Riverine Haplochromine Cichlids. By Cyprian Katongo

This study aimed at reconstructing the phylogenetic relationships among Zambian riverine haplochromine cichlids and at reviewing their taxonomic status. The riverine haplochromines used were from the Zambezi River and the Luapula-Congo River systems. The target taxa were from two major groups: the Serranochromine species and the *Pseudocrenilabrus* species complex. Phylogenetic studies were based on mitochondrial DNA sequence analysis while the taxonomic studies were based on principal component analysis (PCA) of morphological characters. The results indicate that the Zambian riverine haplochromines share a common ancestor, which is likely to have originated from the Congo drainage and later colonized the Zambezi drainage. The results also indicate a need to revise the genera *Serranochromis, Sargochromis*, *Pseudocrenilabrus* and *Pharyngochromis* from all the rivers of Zambia. There are likely to be a number of undescribed species and/or misidentifications within these genera. There might be at least 3 species of *Pseudocrenilabrus* in Zambia. The *Pseudocrenilabrus* 'sp haplochromis-orange' population from the Lake Mweru / Mwatishi River confluence is morphologically the most distant from the other populations. Some of the *Pseudocrenilabrus* species from Luc de Vos' collection from the Lufubu River drainage, held in the Royal Museum of Central Africa at Tervuren, Belgium (MRAC 96-031-P and MRAC 96-083-P) may be members of a new species. The *Sargochromis* specimens from the Luapula, collected by Van Zwieten (MRAC 9-035-P) and those collected during this study (MRAC A4-25), both held at the Royal Museum of Central Africa, appear to consist of two morpho-types with one of them likely to be *Sargochromis mellandi* while the second is likely to be different. There seem to be two or more undescribed species of *Sargochromis* collected from the Kafue River. One of them, a slender type with a long snout (*Sargochromis* 'sp longsnout-kafue'), was quite abundant during the field trip of the summer of 2002, while another one with longitudinal stripes was less abundant (*Sargochromis* 'sp striped'). Van der Waal's collection (RUSI 062863) has one *Sargochromis* species collected from below the Itezhi tezhi Dam, on the Kafue River. This taxon (*Sargochromis* 'sp zambia') is very likely to be a new record in the Kafue system or even an undescribed species. The *Pharyngochromis* group seems to consist of at least more than one species. The *Pharyngochromis acuticeps* from Lake Kariba fits very well into the description of Skelton (1993) and Greenwood (1992). The other *Pharyngochromis* individuals are generally bigger in size with a different coloration on the head. There seem to be some individuals, which are intermediate between *Sargochromis* and *Pharyngochromis* from the Upper Zambezi. There was one *Serranochromis* 'sp elongate', collected from Lake Mweru, which is morphologically similar to (but does not fit into the description of) either *S. macrocephalus* or *S. stappersi*, which have both been documented in that lake. It has a close mitochondrial phylogenetic relationship with *Serranochromis angusticeps*.

Acknowledgements

I would like to thank Prof. Dr. Christian Sturmbauer of the University of Graz, Austria for encouraging me to take up the Ph.D. studies and for being my supervisor and mentor. Members of his research group were very helpful during this study. The Austrian Exchange Services (OAD) provided the fellowship that enabled me carry out most of the work related to my PhD studies.

The Royal Museum of Central Africa at Tervuren, Belgium provided two travel fellowships that enabled me travel to the museum on two occasions on taxonomic study visits and a grant that partially met the costs of my fieldwork in Zambia. I would like to express my gratitude to Prof. Dr. Jos Snoeks, Curator of Fishes at the Royal Museum and visiting Professor at the Catholic University of Leuven, for supporting my application for the two fellowships and the grant. Prof. Dr. Snoeks was also one of my supervisors during my PhD studies. I will always cherish the brief meeting I had with King Albert II of Belgium when he visited the museum while I was there in the summer of 2004!

I would also like to express my thanks to the University of Zambia for granting me study leave to enable me do this PhD training. Dr. Luke Mumba, my supervisor at the University of Zambia, helped by facilitating these studies and by making helpful comments and suggestions that greatly improved my thesis. Dr. Wanchinga and Dr. Nkunika as successive heads of the Department of Biological Sciences gave me administrative support during the initial stages, during fieldwork and in the final stages of my studies. Professors K. Mwauluka, K.J. Mbata and E.N. Chidumayo gave me helpful advice and suggestions during the formulation of the project. Dr. J. Munyandorero and Mr. H.G. Mudenda provided very useful scientific and strategic advise during the initial stages of the studies.

I would like to thank the Department of Fisheries of the Ministry of Agriculture and Co-operatives, Republic of Zambia, for the logistical support I received during my fieldwork. Special thanks go to Mr. Cyprian Kapasa, Mr.J. Mwango and Mr. Nkhata of the Fisheries Department Headquarters at Chilanga; Mr. Silumesi and Staff of Siavonga Station; Mr. Munyakasa and staff of Lukanga Station; Dr. H. Phiri and Staff of Mpulungu Station; Mr. Chitamuka of Samfya Station; Mr. Chiluba and Staff of Nchelenge Station; Mr. Sitali and Mr. Lusenga of Mambova Station, and Mr. Bright Nyimbili and Mr. Munyenyembe of Sinazongwe Station.

I would also like to thank my wife, Mirriam, and my three children Linda, Maggie and Cyprian Junior, for accepting to live separately from me during the period I was pursuing part of my studies in Austria and Belgium.

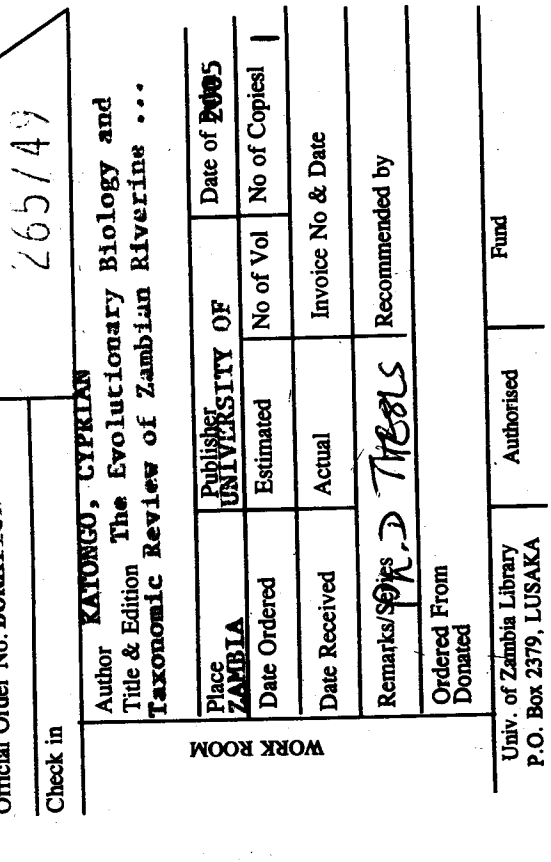
I would finally like to mention my appreciation of the role played by my late father, Mr. Cyprian Walobele who enrolled me into school and kept encouraging me to persevere.

Table of Contents **Page Number**

List of Figures ----------Page Number

List of Tables

Table number **Page**



List of Plates

Plate **Page Number**

List of symbols and abbreviations

bp – Base pairs

CTR – Chain Termination Reaction

Cyt B – Cytochrome B

DNA – Deoxyribonucleic Acid

ddNTPs – Dideoxy Nucleotide Triphosphates

dNTPs – Deoxy Nucleotide Triphosphates

Dloop – Loop in the Control Region of the Mitochondrial DNA

ExoSAP-IT – Trade Mark for PCR Product Purification kit

H 15 149 – DNA Primer for the heavy chain of Cytochrome B

K^+ - Potassium ions

L 14 124 – DNA Primer for the light chain of Cytochrome B

L-ProF – DNA Primer starting at the Proline end of the D-loop

Met – DNA Primer starting at the Methionine end of the NADH gene

mg – Milligrams

Mg^{2+} - Magnesium ions

ml - Millilitre

MZR – Middle Zambezi River

ND2.2A – DNA Internal Primer for the NADH gene

NADH2 /ND2 – Nicotinamide Adenine Dinucleotide Hydrogenated Sub-unit 2

PCA – Principal Component Analysis

PCR - Polymerase Chain Reaction

RFLP – Restriction Fragment Length Polymorphism

RNA – Ribonucleic Acid

rpm – Revolutions per minute

Thr2 – DNA Primer starting at the Threonine end of the D-loop

Trp – DNA Primer starting at the Tryptophan end of the NADH gene

TRM – termination Reaction Mix

UZR – Upper Zambezi River

4 – CLMA – Four-Cluster Likelihood Mapping Analysis

°C – Degrees Celsius

Preface

The riverine haplochromine group of cichlids is closely related to the tilapias but has not been as widely studied as the tilapias, although it has a number of species with a lot of potential for aquarium and aquaculture use. The use of a combination of morphological and molecular methods has helped resolve a number of outstanding evolutionary and taxonomic questions within the target group and its hoped that this approach will give the reader a comprehensive set of diagnostic tools for similar systematic studies. This work could not be exhaustive because of time and material constraints. The results suggest the possible route of colonisation of Zambian rivers by the riverine haplochromines. The need to review the taxonomic status of the riverine haplochromines is evident from the results of this study.

Chapter 1 - Introduction and Background

1.1. Statement of the problem

While the biodiversity of fishes from the Great East African Lakes; Victoria, Malawi and Tanganyika has been studied extensively (Fryer and Iles, 1972; Poll, 1986; Skelton, 1988, 1994; Brichard, 1989; Meyer *et al.*, 1990; Meyer, 1993a, 1993b; Rossiter, 1995; Ruber *et al.*, 1997, 1999; Salzburger *et al.*, 2002; Sturmbauer and Meyer, 1992,1993; Sturmbauer *et al.*, 1994, 1997; Sturmbauer, 1998; Kohda *et al.*, 1996; and Verheyen *et al.*, 1996), much less is known about the ichthyofauna of South Eastern African rivers. Since Zambian river basins have not yet been as systematically surveyed as Lake Tanganyika, the full range of the diversity of riverine cichlid species (in general and riverine, haplochromine cichlid species in particular) is unknown. Due to this fact, it is expected that several riverine cichlid species still await their discovery and this poses taxonomic challenges.

Bell-Cross (1965a) has suggested that the Upper Zambezi/Okavango and Kafue systems formed part of the late Tertiary western drainage basin whose fish fauna showed affinities with the Kasai and Lualaba sub-systems of the Congo River. He considered the Middle/Lower Zambezi to be part of an eastern drainage basin isolated from its western counterpart and whose fauna was considered to be mixed; partly of Nilotic and partly of Lualaba /Congo origin. Most of the current taxonomic knowledge on Zambian Cichlids stems from the relatively old work of the sixties and seventies (Jackson, 1961; Bell-Cross, 1965a, 1965b, 1966, 1972, 1975 and 1976; Jubb, 1964, 1967; Gaigher *et al.*, 1973; Kenmuir, 1983) and none of these studies were placed in a phylogenetic context. To date, only an incomplete coverage of the Zambian Cichlid fauna is found in regional fish guides (Simbotwe and Mubamba, 1993; Skelton, 1991, 1993, 1998, 2001). There is no local reference collection of Type specimens of riverine haplochromine species in Zambia and no adequate identification tools in the form of an identification key.

The morphological differences separating closely related cichlid species are often very small and there is a relatively great intraspecific variation of ranges in the mean values or modes for most morphological characters (Barel *et al.*, 1977; Greenwood, 1984). These small differences often concern squamation patterns and male breeding coloration. It is thus very difficult to construct a simple identification-key for haplochromine cichlids. This leads to the use of a number of taxonomic characters to identify a specimen with certainty. For a definitive identification, comparison with the museum's reference collection is required (Barel *et al.*, 1977).

With the advent of molecular phylogenetic techniques, the taxonomic status of the riverine haplochromine cichlids can be revised by combining morphological and molecular data. Molecular biological data, which provide a new source for constructing phylogenetic trees, are able to reveal differences even among taxa, which are impossible to differentiate on the basis of morphological characters (Sturmbauer and Meyer, 1992). However, the phylogeny of haplochromine cichlid fishes, based solely on molecular data is not always reliable because phylogenetic trees based on single genes might sometimes not correspond to the species tree, so that incorrect phylogenetic hypotheses might be derived. Moreover, inadequate sampling of representatives of the taxa in question might also lead to incorrect hypotheses, when samples do not represent the entire distribution range. Therefore, this study adopted a combination of molecular and morphological techniques as the most promising strategy to obtain reliable results.

1.2. Objectives of the study

The objectives of this study were to:

1) Develop phylogenetic trees of Zambian riverine haplochromine cichlids based on molecular biological data.

2) Review the taxonomic status of the riverine haplochromines from Zambia based on morphological data.

3) Develop an identification key of Zambian riverine haplochromine cichlids, which is based on both morphological identification and molecular characterization.

4) Establish a scientific reference collection of the haplochromine riverine cichlid fishes from selected river basins of Zambia.

1.3. Significance of the Study

This study aims at adding to the existing knowledge on the Zambian cichlid fauna. Quicke (1997) noted that "Generally speaking, well-founded phylogenetic hypotheses can be used to generate experimentally testable predictions about the evolution of many biological systems from biochemical and physiological to behavioural studies". With respect to the Zambian cichlid fauna it is thus of prime importance to reconstruct the phylogenetic relationships of the riverine cichlids covering their entire distribution area of each species and to establish their biogeographic affinities. This study focuses on two groups belonging to the haplochromine cichlids; the Serranochromine cichlids and the species of the genus *Pseudocrenilabrus*. The combination of comparative morphological and molecular biology techniques is expected to help resolve conflicts about the taxonomic relationships. Another aim of this study is to establish a reference collection in Zambia and to elaborate a revised identification key for Zambian riverine cichlids to aid future work.

1.4. Research questions and hypotheses

With the above background and objectives, the study provided a framework for testing the following hypotheses:

1). There is an evolutionary relationship between the cichlids of Congo River Basin and those of the Zambezi River Basin.

2). The riverine haplochromine cichlids of Zambia first colonized the Luapula (Congo) and later the Zambezi River system.

3) Geological processes altered the Zambian drainage systems thereby influencing the dispersal and evolution of cichlid fishes.

Chapter 2 - Literature Review

2.1. Cichlid fishes

Cichlids are perciform fishes found in fresh or brackish waters of Africa, Syria, India, Ceylon, and America, mostly from tropical Africa and tropical America (Boulenger, 1915; Lowe-McConel, 1988; Stiassny, 1991). The family Cichlidae is a highly diverse taxon of freshwater fishes with an estimate of over 2,000 species. They have a diversity of colours, shapes and sizes and are distributed in regions with water temperature of about 20°C or higher. Cichlids have clearly visible cycloid or ctenoid scales on their bodies, one pair of single nostrils, a lateral line in two separate series on each side of the body and a dorsal fin with continuous spiny and soft rays. They have an advanced parental care system for their young such as mouth brooding and bi-parental care (Jackson, 1961; Brichard, 1989; Skelton, 1993, 2001; Sturmbauer *et al.*, 1993; Snoeks, 1994). In terms of scientific interest, cichlid communities, particularly the species flocks in lakes, have become important model systems for the study of evolution, behaviour, physiology, parental care, mating systems, sexual selection, and functional morphology (Brichard, 1989; Skelton, 1994, 2001). Moreover, several cichlid species are frequently kept in aquaria and many larger species are commercially important sources of animal protein. Studies have also, shown the potential of some cichlid species in the biological control of bilharzia transmitting snails (Chimbari *et al.*, 1986, 1987a, 1987b; Brodersen *et al.*, 2002).

2.2. Speciation in cichlid fishes

Cichlid fishes provide better and more illuminating examples of evolution and speciation than the traditionally cited Darwin's finches (Ribbink, 1988). Speciation and adaptive radiations of cichlids have been more evidently pronounced in lacustrine than in riverine environments, probably facilitated by repeated isolation of sub-populations caused by historical fluctuations of the water level in these lakes (Sturmbauer, 1998).

The center of endemicity of the cichlids is located in the East African Great Lakes (Tanganyika, Malawi and Victoria). These lakes are located in the Rift Valley and they harbor more than two-thirds of the estimated 2,000 species in the family. The species flocks of cichlid fishes in the lakes of East Africa are the most spectacular example of adaptive radiation and ecological plasticity among living vertebrates (Greenwood 1984; Sturmbauer and Meyer, 1992; Kocher, *et al.*, 1993). The Tanganyikan cichlid radiation is the oldest, and the species flocks in lakes Malawi and Victoria are derived from one of the ancient lineages found in Lake Tanganyika (Sturmbauer, 1998; Salzburger *et al.*, 2002). About 90-100 out of the estimated 2,000 cichlid species live in African rivers (Greenwood, 1984), about 18 species live in Madagascar (Farias *et al.*, 2000).

The evolutionary success of the cichlids is mainly explained by two key innovations. The first one lies in the anatomy of the pharyngeal apophysis, providing a second set of jaws functionally decoupled from oral teeth (Greenwood, 1978). Small modifications of this structure allow utilization of new food resources, so that new ecological niches can be rapidly occupied and new species can arise (Meyer, 1987; Greenwood, 1984; Sturmbauer, 1998). The second key innovation is the highly specialized mating and breeding behaviour. Mouth brooding, a character of many cichlids is the most advanced parental care systems known among fishes (Sato, 1986). This suggests that sexual selection is in part responsible for the rapid speciation in cichlids (Sturmbauer and Meyer, 1992).

2.3. Distribution of cichlid fishes

2.3.1. Global biogeographic distribution of the major groups of cichlids

There are three major groupings of cichlids. These are: the Malagasy/Indian cichlids, the African cichlids and the Neotropical cichlids. These groupings reflect the past geological events, which led to the separation of the Gondwana subcontinent. This subcontinent was made up of the present day India, Africa and South America. The first separation is believed to be between the Indian peninsula and the rest of the subcontinent, followed by

the separation of Malagasy from the eastern part of the subcontinent and finally Africa is believed to have separated from the Americas (Bowmaker *et al.*, 1978; Farias *et al.*, 2000). Cichlid evolutionary studies, therefore compliment geologic data used for the explanation of the historical events that led to the present geography of the continents where cichlid fishes are found.

2.3.2. Zoogeographical affinities of Zambian cichlid fishes

The Zambian fauna area consists of two great river drainage systems, the Zambezi and the Zambian Congo (Jackson, 1961; Beadle, 1981). Due to the close association with this major African watershed, the Zambian cichlid fauna holds a key position for understanding the evolutionary history of all African cichlids (Greenwood, 1984; Mayr, 1984; Stiassny, 1991).

The southern boundary of Zambia is the Zambezi River. To the east the mountain scarps separating Zambia from Malawi form a natural barrier to the fish faunas of the two areas. Similarly, just inside the northern boundary the high ground immediately south of Lake Tanganyika is a watershed, which separates the Tanganyika fauna from those of the extreme upper Congo. To the west the Chambeshi River flows to the Bangeweulu swamps, from where the Luapula River connects to Lake Mweru and continues northwards. The rapids on the Luapula after Lake Mweru separate the region known as the Zambian Congo from the remaining Upper Congo (Jackson, 1961).

As far as the distribution of fish is concerned the Okavango, Upper Zambezi and Kafue Rivers form a single zoogeographic region. It has a number of species in common with the southern tributaries of the Congo River, especially the River Kasai and the west-flowing rivers of Angola. Bell-Cross (1972) has suggested that the distribution of the Zambezi fishes can best be explained as the result of a previous separation of the Okavango/Upper Zambezi/Kafue river system from the Middle-Lower Zambezi with consequentual different origins of much of their faunas. Prior to the Pleistocene, it is supposed that the former may have flowed westwards to join the Congo basin and that

their direction was later reversed to connect into the Zambezi. There is some evidence of a recent link between the Upper Zambezi and Kasai Rivers. It thus emerges that the past history of the Zambezian fish fauna is exceptionally difficult to unravel (Bell-Cross 1972).

2.4. Lake Tanganyika cichlid flock and their relationships to the riverine haplochromine fauna

2.4.1. Haplochromine cichlids

Within Southern Africa, in more general terms, there are two main cichlid lineages, the tilapiines which are chiefly plant or sediment feeders, and the haplochromines, which tend to be predators. The young tilapiines have a dark eyespot at the base of the soft dorsal fin, called the tilapia spot. Adult haplochromines usually have a series of coloured spots or ocelli, often called egg-spots or egg dummies on the anal fin (Skelton, 2001).

Riverine haplochromines of Zambia include the genera *Pseudocrenilabrus*, *Thoracochromis, Orthochromis* and the Serranochromine cichlids (Skelton, 1998, 2001). This group of fishes is closely related to the H-lineage of the Tanganyika cichlid flock (Nishida, 1991; Salzburger *et al.*, 2002). Specifically, the Serranochromine cichlid group includes the genera *Sargochromis*, *Serranochromis*, *Pharyngochromis* and *Chetia* (Greenwood, 1993).

2.4.2. Tanganyika cichlid fish radiation

Because of its old age, Lake Tanganyika habours cichlid fishes which have a very close relationship with ancestors of riverine haplochromines of South Eastern Africa (Bell-Cross, 1965a; Fryer and Iles, 1972; Bowmaker *et al.*, 1978; Nishida, 1991; Salzburger *et al.*, 2002; Sturmbauer *et al.*, unpblished).

With an estimated age of 9 to 12 million years old (Salzburger *et al.*, 2002), Lake Tanganyika has a complex geological history, characterised by periods of geological activity, climatic change and extended intervals of substantially lower lake level (Cohen *et al.*1997). The present distribution of many endemic cichlid species has been greatly affected by the fluctuations of the lake levels in the past. These fluctuations caused alternate expansions, contractions and sometimes disappearance of sandy beaches, swamps and rocky coasts (Brichard, 1989; Baric, 2000).

Studies on the adaptive radiation of cichlids in Lake Tanganyika suggest that several lineages arose with dramatic speed immediately after the formation of a lacustrine habitat. This response is manifested in short branch lengths on phylogenetic trees, which make evolutionary reconstruction difficult (Nishida, 1991; Sturmbauer and Meyer, 1993; Kocher *et al.*, 1993).

The cichlid fauna of Lake Tanganyika have been grouped into twelve tribes (Poll, 1986). Eight of these tribes are endemic to the Lake (Rossiter, 1995). A polyphyletic origin of the Lake Tanganyika species flock has been suggested, given that 4 out of the 12 tribes have members that are found outside the lake (Poll, 1986; Nishida, 1991). Neither the geographic origin nor the identity of the relatives outside the lake is known for most endemic mouth brooding lineages in Lake Tanganyika. The relationships of the substrate spawning Lamprologini to the seven mouth brooding tribes: the Eretmodini, Perissodini, Ectodini, Cyprochromini, Limnochromini, Haplochromini and Tropheini have remained unclear, (Salzburger *et. al.*, 2002). Investigators have so far assumed that several endemic mouth brooding tribes originated from riverine colonisers (Fryer and Iles, 1972; Nishida, 1991).

The study of Salzburger *et al.* (2002) was the first to include several riverine haplochromines of the genera *Orthochromis*, *Schwetzochromis*, *Cyclopharynx*, *Pseudocrenilabrus*, *Astatotilapia* and *Astatoreochromis* (Poll, 1986), all of which are potential sister groups to lake endemics. The inclusion of these taxa allowed Salzburger *et al.* (2002) to test the hypothesis that haplochromine colonisers would have seeded the

primary radiation of most mouth brooding lineages in Lake Tanganyika. Two cladogenetic events affecting the Lake Tanganyika cichlid species assemblage were identified. The first series involving the lineages that seeded the Lake Tanganyika radiation is likely to have taken place at an early stage of the lake formation, when the protolakes were quite low (Tiercelin and Mondeguer, 1991). The second series of cladogenetic events appears to represent the primary radiation, taking place in the lacustrine habitat. Salzburger *et al.* (2002) suggest that eight distinct lineages have seeded the Tanganyika radiation: the Tylochromini, two lineages of the Tilapiini, the Bathybatini, the Trematocarini, the Eretmodini, the ancestors of the Lamprologini and the ancestors of the H-lineage (Nishida, 1991), excluding the Eretmodini. The Tanganyikan representatives of the Tylochromini and Tilapiini were either already part of the ancient fauna of the proto-Malagarazi-Congo River before the lake basin formed or entered the lacustrine habitat more recently. The remaining five seeding lineages radiated in step with the development of the current truly lacustrine habitat. Nishida (1991) identified the Trematocarini and the Bathybatini as seeding lineages. Because they both underwent radiation in Lake Tanganyika and no relatives are known in extant river faunas, these endemics may be the only surviving descendants of the two ancient African lineages. The same may be true for the Eretmodini and Lamprologini, which consistently branched before the onset of the primary lacustrine radiation, and for the ancestor of the H-lineage (excluding the Eretmodini).

On the basis of morphological characteristics, Fryer and Iles (1972) suggested that the endemic mouthbrooding cichlid tribes of Lake Tanganyika have originated from riverine haplochromine colonisers. The morphological characteristics regarded as the most important factors in promoting coexistence in cichlids include: dental and pharyngeal specializations of the jaw morphology, as well as dietary specialization (Fryer and Iles, 1972).

Sturmbauer (1998) suggests that when Lake Tanganyika emerged, the potential colonisers were from the surrounding rivers. He notes that the great mobility of riverine fishes results in high gene flow, resulting into sympatric modes of speciation. After the

incipient species become specialized to discontinuous habitats, allopatric modes of speciation play a larger role within the single lake basin. With increasing age, the species become ecologically and morphologically specialized, so that their bonds to particular habitats increase, as their dispersal ability decreases. He suggests that the primary cichlid radiation of the lake was seeded by riverine species, that the formation of the proto-lakes induced the second radiation and finally that the formation of a truly lacustrine environment induced the third radiation. The recent lake level fluctuations have triggered further secondary radiations within the third radiation (Sturmbauer, 1998).

So far the hypothesis that ancestral haplochromines have seeded the primary radiation of the majority of the mouth brooding lineages of Lake Tanganyika remains untested (Salzburger *et al.*, 2002).

2.4.3. Lake Tanganyika cichlid flock and the African haplochromine fauna

Among the eight endemic cichlid tribes of Lake Tanganyika, the tribe Haplochromini, with about 1, 700 species, represents the most species-rich tribe. Some species of the Haplochromini occur in the lake and in surrounding rivers and swamps (Poll, 1986; Salzburger *et al.*, 2002). The tribe Lamprologini, with 84 described species, is the most species-rich Tanganyikan taxonomic lineage and also contains five riverine species found in the Congo River system (Salzburger *et al.*, 2002). Given that the most ancestral splits in the Lamprologini are formed by lake endemics, Sturmbauer *et al.*, (1994) suggest that the radiation of the Lamprologini started in the Tanganyikan basin before the ecosystem was hydrologically and ecologically closed, whereas the riverine species left the lake secondarily.

The Lake Tanganyika cichlid species assemblage has so far been viewed as an evolutionary reservoir of old African lineages (Nishida, 1991). Alternatively it may be viewed as an evolutionary hotspot from which modern African cichlid lineages arose, spreading subsequently to rivers and lakes in other African regions. Part of the

Tanganyikan cichlid species flock is derived from an ancient fauna that no longer exists anywhere else in Africa (Sturmbauer *et al.*, 1994).

2.4.4. The H-lineage of the Tanganyika cichlid flock

Nishida (1991) combined the seven tribes Eretmodini, Limnochromini, Perissodini, Ectodini, Cyprochromini, Haplochromini, and Tropheini into the H-linaege pointing to the possibility that they evolved within the lake from the same riverine ancestor or ancestors. The study of Sturmbauer and Meyer (1993), which included the seven mouth brooding tribes (the H-lineage), suggested a sister group relationship between the haplochromine cichlid *Astatotilapia burtoni* and the endemic Tanganyikan tribe Tropheini. However, this representative was not resolved as the most ancestral branch of the H-lineage. In the phylogeny of Kocher *et al.* (1993), based on mitochondrial NADH dehydrogenase subunit2 gene (ND2), the mouth brooding tribe Eretmodini was placed outside the H-lineage as sister group to the substrate spawning Lamprologini. Five species of Haplochromini from Lake Malawi were resolved as sister groups to the Tropheini. Salzburger *et al.* (2002) identified two Congo River haplochromines as members of the H-Lineage. Further studies placed these species as the sister group of the serranochromine cichlids (Klett and Meyer, 2002; Sturmbauer *et al.*, unpublished).

2.5 Taxonomic challenges in cichlid fishes

Fish taxonomy is a branch of Ichthyology, which deals with the identification of fish and their classification according to presumed relationships. Rules of nomenclature are followed as laid down by the International Code of Zoological Nomenclature (Froese and Pauly, 2000). In this system, a name is valid when it is first published only if it is accompanied by a description and diagnosis, which separates it from related forms at the same level of classification. When a fish is being described, a type specimen should be provided and deposited in a recognised museum for future reference. The specimens on which the original description of a species is based are called types or syntypes. The single specimen on which the original description is made is called the holotype and all

the other specimens of the type series are called paratypes. Paratypes are used together with the holotype in drawing up the original description of the new species. A lectotype is a specimen selected by a later worker from the syntypes to resolve a taxonomic problem and the remaining specimens from the syntype series are referred to as paralectotypes. A single specimen designated as the name bearing type of a nominal species or subspecies for which no holotype, lectotype or syntype is believed to exist, is referred to as a neotype. When groups are revised, taxonomic keys are usually provided that are used to identify fish from a given water body or region (Barel *et al.*, 1977).

Due to their external and anatomic similarity, the taxonomy of the Victoria haplochromine cichlids, for example, is beset with difficulties, and consequently many groups are studied by very few or only one specialist at a time, Boulenger, Regan, Trewavas, and Greenwood being the main researchers focusing on haplochromines in successive generations (Barel *et al.*, 1977).

The evolutionary age of the Tanganyikan species flock is manifested not only in the unparalled degree of morphological and ecological divergence within and among the endemic lineages but also by the absence of morphologically intermediate species that might reveal how the lineages are interrelated. Diversification has been achieved mainly by allometric changes of relatively few morphological structures, rather than true morphological innovations (Greenwood, 1984; Stiassny, 1991) thus posing considerable problems for morphology-based phylogenetic analyses. Within a group of haplochromines, the morphological characters separating species show a high level of individual variability (Greenwood, 1984).

The classification of the numerous members of the family Cichlidae presents great difficulties, and the division into genera is unsatisfactory and open to criticism. The dentition, an important taxonomic character, is for example, subject to variation according to age, or even of a purely individual nature in some species (Snoeks, personal comm.). The total range of external morphology into which many of the haplochromine species fit is extremely narrow and consequently, the morphological differences

separating closely related species is small. Further, haplochromine species demonstrate a relatively great intraspecific variation for most characters. The overall similarity is typified by the fact that, for example, local fishermen of Lake Victoria, who are otherwise very capable in distinguishing species, collectively refer to all haplochromine species by one name – Furu (Barel *et al.*, 1977). However, Ambali *et al.*, 2001 have demonstrated that the local fishermen of Lake Malawi have an elaborate system of naming the different haplochromine species. The commercially important species tend to have fewer names over many different localities of the lake while less important species have different names in the different localities.

The practical difficulties frustrating taxonomic studies on haplochromine cichlids, are related to the time consuming process of familiarizing oneself with the species described and the way in which new species are recognized. There is need to study the reference collections in musea and other research institutions. Although manuals are available, the finer technical details of the taxonomic analysis have been handed down orally between the specialists of successive generations. A simple key to cichlid genera is not always practical as many pertinent features are internal, skeletal or behavioural. Even when the mean values (or modes) of a certain character are clearly separated, there is often a certain overlap of ranges. It is therefore nearly always necessary to register a number of taxonomic characters to identify a specimen with certainty, thus making identification a time-consuming affair. For a final identification, comparison with the museum's reference collection often remains necessary as it is difficult, if not impossible, to keep a clear mental picture of all relevant species in other groups unbiased by the group on which one is currently working (*Barel et al.*, 1977; Skelton, 1993).

2.5.1. Cladistical analysis of cichlids based on morphological characters

Cichlids vary tremendously in shape, yet this variation is largely due to allometric differences rather than to *de novo* evolution of characters. The evolution of cichlids often proceeds through evolution via concertina (Stiassny, 1991). Hence, few morphological characters are available for cladistical analyses of cichlid relationships, and homoplasies

seem to be commonplace. Cladistical analysis based on morphological characters is a difficult exercise, with a lot of pitfalls, nothing compared to the straightforward coding in molecular phylogenetic analysis (Snoeks, Personal communication). Morphological characters have to be coded for them to be used in cladistical analysis (Wiley *et al.*, 1991). In general, characters that can be coded for use in cladistical analyses include morphological, molecular and behavioural characters, as long as they are quantifiable and consistent to each other. However, there is considerable controversy about the justification of combining different data sets in a "total evidence approach" (Sturmbauer, Personal communication).

2.6 Phylogenetic inference in cichlids on the basis of molecular data

Hennig (1965, 1966) was the first to formalize the concept of phylogenetic systematics. Phylogenetic systematics based on molecular data has revolutionized approaches to the biological classification of organisms because it uses data that are independent of morphology. Molecular trees thus provide important feedback to morphology-derived hypotheses. The relations between organisms inferred from molecular sequences usually support those inferred from morphological characters (Patterson *et al.*, 1993; Farias *et al.*, 2000). Most of the discrepancies involve species with morphological differences that are few in number, inconsistent, or inconclusive. A good example is the ancestral relationship between the gorilla, chimpanzee and human. Unresolved on morphological grounds, molecular data strongly favour the gorilla as the first branch off the common ancestor, but only after using 30, 000 of aligned sequences, was statistical significance reached (Hartl, 2000).

Inferring a phylogeny is an estimation procedure. This involves making the best estimate of an evolutionary history based on the incomplete information contained in the data. Since it is possible to postulate evolutionary scenarios by which any chosen phylogeny could have produced the observed data, it is important to have some basis for selecting one or more preferred trees from among a set of possible phylogenies. Phylogenetic inference methods seek to accomplish this in two ways: (1) by defining a specific

sequence of steps (an algorithm) that leads to the determination of a tree; or (2) by defining a criterion for comparing alternative phylogenies to one another and deciding which one is to be preferred according to particular optimality criteria (Hillis *et al.*, 1996).

Purely algorithmic methods combine tree inference and the definition of the preferred tree into a single statement. These methods include all forms of pair-group cluster analysis (e.g. UPGMA) and some other distance methods such as neighbour joining (Saitou and Nei, 1987). The methods tend to be computationally fast because they proceed directly toward the final solution without requiring evaluation of large numbers of competing trees (Hillis *et al.*, 1996). The second class of methods has two logical steps. The first step is to define an optimality criterion for evaluating a given tree – i.e. a score is assigned and subsequently used for comparing one tree to another. The second is to use specific algorithms for computing the value of the objective function and finding the trees that have the best values according to this criterion (a maximum or minimum value, as appropriate). Thus, the evolutionary assumptions made in the first step are decoupled from the computer science of the second step. These methods tend to be slower than those of the first class, a consequence of having to search for the tree(s) with the best score (Hillis *et al.*, 1996).

Most of the analytical techniques result in the inference of an unrooted tree, a phylogeny in which the earliest point in time (the location of a common ancestor) is not identified. The components of a phylogenetic tree go by a variety of names. The contemporary taxa correspond to terminal nodes or tips, also called leaves or external nodes. The branch points within a tree are called internal nodes. Nodes are also called vertices or points. The branches connecting (incident to) pairs of nodes are called edges, links or segments. The peripheral branches refer to branches that end at a tip and interior branches (or in the case of a tree with four terminal nodes, central branch) refer to branches that are not incident to the tip. If just three branches connect to an internal node, then the node represents a bifurcation, or dichotomy. If there are more than three branches connected to an internal node, then the node represents a multifurcation, or polytomy. A tree in which all internal

nodes represent bifurcations is said to be binary, fully resolved, or strictly bifurcating. A tree that contains a single internal node is called a star tree (Wiley *et al.*, 1991).

Experimental characters generally fall into one of two broad categories: discrete characters, and similarities or distances. Characters are also assumed to be homologous, meaning that the character must be defined in such a way that all the states observed over taxa for that particular character must have been derived perhaps with modification, from a corresponding state observed in the common ancestor of those taxa. When dealing with relationships among species rather than genes, the definition is restricted to include only orthologous, as opposed to paralogous or xenologous genes. In general, character data are either qualitative, in which case the possible character states are two or more discrete values; or quantitative, in which case the characters vary continuously and are measured on an internal scale. Qualitative data may be further divided into binary (two possible character states) and multistate (three or more possible character states). Binary data typically represent the presence or absence of some character such as the recognition sequence for a restriction sequence for a restriction endonuclease at a certain map location (restriction site) or a particular allele at an isozyme locus. Multistate characters may be ordered or unordered, depending on whether an ordering relationship is imposed upon the possible character states. Nucleotide sequence data are generally treated as unordered multistate characters, since there is no *a priori* reason to assume, for instance that the state C is intermediate between A and G (Hillis *et al.*, 1996).

Character order defines the allowed character-state transformations, whereas character polarity refers to the direction of character evolution. Estimation of character polarity generally involves an assessment of the observed character state most likely to represent the ancestral condition (i.e. the state found in the most recent most common ancestor of the taxa under study) (Mabee, 1989).

From the origin of natural sciences, many biologists tried to categorise the organismic diversity, without believing in evolution. This view of life began to change with Charles Darwin and Alfred R. Wallace's theory of natural selection. This theory formed the basis

for modern evolutionary thinking. Since the early 1950's, objective criteria for reconstruction of the evolutionary history based on shared characters of living and extinct organisms were formulated. In the 1960's these approaches were used to develop the first computer programs for phylogenetic estimations. Nearly at the same time methods for clarification of the molecular structure of proteins and DNA were introduced in evolutionary biology, so that a great number of discrete characters could be used for phylogenetic analysis. Due to the development of new molecular biological approaches in addition, or in contrast, to the earlier exclusively used anatomical, morphological and ecological data, the possibility of analysing huge data sets on computers dramatically enhanced the possibilities for evolutionary biologists. For the past few decades, biologists from many different disciplines have turned to phylogenetic analyses to interpret variation in biological systems. Today molecular investigations of evolutionary history are being used to study questions as diverse as epidemiology of AIDS, origin of life, etc. (Hillis *et al.*, 1996).

2.7. Molecular Clock

The hypothesis of the molecular clock is that the rate of molecular change within genes is similar enough, to make phylogenetic predictions possible (Zuckerkandl and Pauling, 1962, 1965). Today this theory about a rate constancy of molecular changes is obsolete, because many results have been obtained that do not correlate with paleontological findings. Additionally, heterogeneity of evolutionary rate of different nucleotides, genes and genomes could be shown (Li & Graur, 1991). Nevertheless, there are adequate molecular clocks for determined genes within a group of organisms. For this reason it is possible to reconstruct phylogenetic trees without any failure, if the base substitution rates of the sequences of a particular dataset lie within 0.5% significance interval. The significance borders of the base substitution rates of closely related species comply with a Poisson distribution, if assumed that molecular change is a linear function of time (Hillis *et al.*, 1996). To test the hypothesis of the constant molecular clock within a dataset and to identify taxa with significantly faster or slower base substitution rates (99% significance interval of the z-value) a branch length test using LINTRE (Takezaki *et al.*,

1995) can be used. This test shows the deviation of the root-to-tip-distances (genetic distance from the base to the tip of the phylogenetic tree) of individual taxa from the mean.

Inspite of variation in rates of evolution among different molecules, the average rate of molecular evolution often manifests an appropriate uniformity throughout long periods of evolutionary time. A common example of clocklike uniformity in amino acid substitutions is illustrated in the gene tree of the α-globin gene among different animal taxa.

Takezaki and co-workers (1995) developed a method of constructing a linearized tree under the assumption of a molecular clock. It consists of two tests of the molecular clock for a given topology: the two-cluster test and branch-length test. The two-cluster test examines the hypothesis of the molecular clock for the two lineages created by an interior node of the tree, where as the branch length test examines the deviation of the branch length between the tree root and a tip from the average length. Sequences evolving excessively fast or slow at a high significance level may be eliminated. A linearized tree is then constructed for a given topology for the remaining sequences under the assumption of rate constancy (Takezaki *et al.*, 1995).

Aging of cichlid taxa is based on comparisons of age estimate made using geological data of the area or water body where the cichlids live. This is also done by comparing the age estimate of one taxon with another taxon whose own age estimate is based on geological data. This is because there are no fossils of cichlids to be used to establish the age structure of the cichlids (Duftner, 2003, Personal communication).

Chapter 3 Materials and Methods

3.0. Research design

3.1. Sampling

Sampling of the fishes was carried out in close cooperation with the Department of Fisheries, Ministry of Agriculture and Cooperatives, Republic of Zambia, which has established several monitoring stations on all important water bodies of Zambia. Fish samples were collected from the Zambezi and the Luapula River Systems (Figures 7 and 8). The sites were selected at random to represent the various parts of the Zambezi and Luapula systems. Sampling was through the use of a series of gill nets and with the help of the local fishermen and fish traders. Samples from fishermen and traders were chosen according to the target species and five specimens were collected for each species and for each locality. Each specimen collected was measured and a colour picture taken.

A fin clip was subsequently taken from the right pectoral fin and stored in 90% ethanol for genetic analysis. The specimens (whole fish) were transferred to buffered formalin (10%) for identification and morphological analysis. Finally, selected voucher specimens from the samples were used to set up a reference collection in the Zoological Museum of the Biological Sciences Department, University of Zambia with replicates deposited in the Royal Museum of Central Africa at Tervuren, Belgium. The list of specimens used in both morphological and molecular analyses is given in Table1. The sample list with additional (museum) specimens used in this study is given in Appendix 2 (Table14) while additional DNA sequences used in this study are given in Appendix 3 (Table 15).

Table 1. List of core specimens used in both morphological and molecular analyses. These are the samples collected during fieldwork

Morph. No	DNA no.	Collection no.	Identity	Locality
CK156b	2559	UZ4F7	*Serranochromis* 'sp elongate'	Mifimbo Lake Mweru
CK157b	2560	UZ4F8	*Serranochromis* 'sp elongate'	Mifimbo Lake Mweru
CK158	2561	UZ4F9	*Serranochromis* 'sp elongate'	Mifimbo Lake Mweru
CK159	2562	UZ4F10	*Serranochromis* 'sp elongate'	Mifimbo Lake Mweru
CK160	2563	UZ4G1	*Serranochromis* 'sp elongate'	Mifimbo Lake Mweru
CK161	-	UZ4G8	*Serranochromis* 'sp elongate'	Mifimbo Lake Mweru
CK162	-	UZ4G9	*Serranochromis* 'sp elongate'	Mifimbo Lake Mweru
CK163	-	UZ4G10	*Serranochromis* 'sp elongate'	Mifimbo Lake Mweru
CK164	-	UZ4H1	*Serranochromis* 'sp elongate'	Mifimbo Lake Mweru
CK165	-	UZ4H2	*Serranochromis* 'sp elongate'	Mifimbo Lake Mweru
CK166	2738	UZ4H7	*Serranochromis stappersi*	Mifimbo Lake Mweru
CK167	2739	UZ4H8	*Serranochromis stappersi*	Mifimbo Lake Mweru
CK168	2740	UZ4H9	*Serranochromis stappersi*	Mifimbo Lake Mweru
CK169	2741	UZ4H10	*Serranochromis stappersi*	Mifimbo Lake Mweru
CK170	2742	UZ4I1	*Serranochromis stappersi*	Mifimbo Lake Mweru
CK171	-	UZ4I2	*Serranochromis stappersi*	Mifimbo Lake Mweru
CK172	-	UZ4I3	*Serranochromis stappersi*	Mifimbo Lake Mweru
CK173	-	UZ4I4	*Serranochromis stappersi*	Mifimbo Lake Mweru
CK174	2471	UZ4E2	*Serranochromis stappersi*	Mifimbo Lake Mweru
CK175	2472	UZ4E3	*Serranochromis stappersi*	Mifimbo Lake Mweru
CK176	2734	UZ4E4	*Serranochromis stappersi*	Mifimbo Lake Mweru
CK177	2735	UZ4E5	*Serranochromis stappersi*	Mifimbo Lake Mweru
CK178	2736	UZ4E6	*Serranochromis macrocephalus?*	Mifimbo Lake Mweru
CK179	2634	UZ4B9	*Serranochromis robustus*	Chalwe Lake Bangweulu
CK180	2635	UZ4B10	*Serranochromis robustus*	Chalwe Lake Bangweulu
CK181	2477	UZ4C9	*Serranochromis robustus*	Lwaka Lake Bangweulu
CK182	2478	UZ4D2	*Serranochromis robustus*	Kapapa Lake Bangweulu
CK183	2636	UZ4D3	*Serranochromis robustus*	Kapapa Lake Bangweulu
CK184	-	UZ4D4	*Serranochromis angusticeps*	Kapapa Lake Bangweulu
CK185	-	UZ4D5	*Serranochromis angusticeps*	Kapapa Lake Bangweulu
CK186	-	UZ4D8	*Serranochromis angusticeps*	Chalwe Lake Bangweulu
CK187	2607	UZ4D9	*Serranochromis angusticeps*	Chalwe Lake Bangweulu
CK188	2608	UZ4A7	*Sargochromis mellandi*	Lwaka Lake Bangweulu
CK189	2475	UZ4A8	*Sargochromis mellandi*	Chalwe Lake Bangweulu
CK190	2475	UZ4A8b	*Sargochromis mellandi*	Chalwe Lake Bangweulu
CK191	2476	UZ4C10	*Sargochromis mellandi*	Musanse Lake Bangweulu
CK192	2476	UZ4D1	*Sargochromis mellandi*	Musanse Lake Bangweulu
CK193	2476	UZ4D6	*Sargochromis mellandi*	Musanse Lake Bangweulu
CK194	2670	UZ4D7	*Sargochromis mellandi*	Musanse Lake Bangweulu
CK195	2671	UZ4E7	*Sargochromis mellandi*	Mifimbo Lake Mweru
CK188	-	UZ4A7	*Sargochromis mellandi*	Mifimbo Lake Mweru
CK196	-	UZ4E8	*Sargochromis mellandi*	Mifimbo Lake Mweru

Table 1 continued

CK197	2473	UZ4E9	*Sargochromis mellandi*	Mifimbo Lake Mweru
CK198	2473	UZ4E10	*Sargochromis mellandi*	Mifimbo Lake Mweru
CK199	2474	UZ4F1	*Sargochromis mellandi*	Mifimbo Lake Mweru
CK200	2474	UZ1A2	*Serranochromis macrocephalus*	Kafue Market Kafue River
CK201	2672	UZ1A3	*Serranochromis macrocephalus*	Kafue Market Kafue River
CK202	2673	UZ1B4	*Serranochromis macrocephalus*	Kafue Bridge Kafue River
CK203	-	UZ1B6	*Serranochromis macrocephalus*	Kafue Bridge Kafue River
CK204	2556	UZ1F"	*Serranochromis macrocephalus*	Lukanga Swamps Kafue er
CK205	2557	UZ1D8	*Serranochromis macrocephalus*	Lukanga Swamps Kafue River
CK206	-	UZ1E/ CK391	*Serranochromis macrocephalus*	Lukanga Swamps Kafue River
CK207	2215	UZ1G7	*Serranochromis macrocephalus*	Mazabuka Kafue River
CK208	-	UZ1G8	*Serranochromis macrocephalus*	Mazabuka Kafue River
CK209	2264	UZ3D4	*Serranochromis macrocephalus*	Mongu Market UZR
CK210	2553	UZ3D5	*Serranochromis macrocephalus*	Libonda Camp UZR
CK211	2276	UZ3D6	*Serranochromis macrocephalus*	Mongu Market UZR
CK212	-	UZ3D7	*Serranochromis macrocephalus*	Mongu Market UZR
CK213	2555	UZ3D8	*Serranochromis macrocephalus*	Mongu Market UZR
CK214	2528	UZ1A1	*Serranochromis angusticeps*	Chanyanya Kafue River
CK215	2529	UZ1A7	*Serranochromis angusticeps*	Chanyanya Kafue River
CK216	2530	UZ1A8	*Serranochromis angusticeps*	Chanyanya Kafue River
CK217	2531	UZ1B3	*Serranochromis angusticeps*	Kafue Bridge Kafue River
CK218	2532	UZ1D1	*Serranochromis angusticeps*	Lukanga Swamps Kafue River
CK219	2593	UZ1D3	*Serranochromis angusticeps*	Lukanga Swamps Kafue River
CK220	2260-2592	UZ1D2	*Serranochromis angusticeps*	Lukanga Swamps Kafue River
CK221	2594	UZ1D9	*Serranochromis angusticeps*	Lukanga Swamps Kafue River
CK222	-	1A3?	*Serranochromis angusticeps*	Chanyanya Kafue River
CK223	2202-2637	UZ1A5	*Serranochromis angusticeps*	Kafue Bridge Kafue River
CK224	-	UZ3C3	*Serranochromis angusticeps*	Mongu Market UZR
CK225	-	UZ3C2	*Serranochromis angusticeps*	Mongu Market UZR
CK226	2596	UZ3C1	*Serranochromis angusticeps*	Mongu Market UZR
CK227	-	UZ3C4	*Serranochromis angusticeps*	Mongu Market UZR
CK228	-	UZ3G5	*Serranochromis angusticeps*	Mongu Market UZR
CK229	2612	See CK231	*Serranochromis altus?*	Mazabuka Kafue River
CK230	2680	UZ3H6	*Serranochromis altus*	Nalyele Mongu UZR
CK231	2612	UZ1G6/ CK229	*Serranochromis angusticeps*	Mazabuka Kafue River
CK232	2614	UZ1A6	*Serranochromis robustus*	Chanyanya Kafue River
CK233	2285-2615	UZ1F6	*Serranochromis robustus*	Lukanga Swamps Kafue River
CK234	2616	UZ1G7	*Serranochromis robustus*	Mazabuka Kafue River
CK235	2617	UZ1H4	*Serranochromis robustus*	Mazabuka Kafue River
CK236	2618	UZ1H5	*Serranochromis robustus*	Mazabuka Kafue River
CK237	-	UZ1Hi	*Serranochromis robustus*	Mazabuka Kafue River
CK238	-	UZ1Hii	*Serranochromis robustus*	Mazabuka Kafue River
CK239		UZ3A2	*Serranochromis robustus*	Mukakani Camp UZR

Table 1 continued

CK240	-	UZ3A3	*Serranochromis robustus*	Mukakani Camp UZR
CK241	2625	UZ3B7	*Serranochromis robustus*	Mongu Market UZR
CK242	2621	UZ3B3	*Serranochromis robustus*	Mongu Market UZR
CK243	2622	UZ3B4	*Serranochromis robustus*	Mongu Market UZR
CK244	2623	UZ3B5	*Serranochromis robustus*	Mongu Market UZR
CK245	2624	UZ3B6	*Serranochromis robustus*	Mongu Market UZR
CK246	2627	UZ3E1	*Serranochromis robustus*	Likundu Camp UZR
CK247	2628	UZ3E2	*Serranochromis robustus*	Likundu Camp UZR
CK248	-	UZ3E3	*Serranochromis robustus*	Likundu Camp UZR
CK249	2367	UZ3H8	*Serranochromis species*	Mukakani Camp UZR
CK250	2368	UZ3H9	*Serranochromis species*	Mukakani Camp UZR
CK251	2203-2638	UZ1A10	*Sargochromis species*	Chanyanya Kafue River
CK252	2202-2637	UZ1A5	*Sargochromis species*	Chanyanya Kafue River
CK253	2286	See CK258/CK284	*Sargochromis carlottae?*	Lukanga Swamps Kafue River
CK254	2669	UZ1F1	*Sargochromis carlottae?*	Lukanga Swamps Kafue River
CK255	-	UZIF2	*Ctenopoma species*	Lukanga Swamps Kafue River
CK256	-	See CK271	*Sargochromis species*	Lukanga Swamps Kafue River
CK257	-	CK257	*Sargochromis species*	Lukanga Swamps Kafue River
CK258	2286	UZ1E6/CK284	*Sargochromis carlottae?*	Lukanga Swamps Kafue River
CK259	2207-2639	UZ1C1	*Sargochromis species*	Chanyanya Kafue River
CK260	2206	UZ1B10	*Sargochromis species*	Chanyanya Kafue River
CK261	2205	UZ1B9	*Sargochromis species*	Chanyanya Kafue River
CK262	2204	UZ1B8	*Sargochromis species*	Chanyanya Kafue River
CK263	2201	UZ1A4	*Sargochromis species*	Chanyanya Kafue River
CK264	2271	UZ1B1	*Sargochromis species*	Chanyanya Kafue River
CK265	2261-2641	UZID5	*Sargochromis carlottae?*	Lukanga Swamps Kafue River
CK266	2642	UZ1D6	*Sargochromis carlottae?*	Lukanga Swamps Kafue River
CK267	2208-2640	UZ1C2	*Sargochromis carlottae?*	Lukanga Swamps Kafue River
CK268	2287-2643	UZ1F7	*Sargochromis carlottae?*	Lukanga Swamps Kafue River
CK269	2644	UZ1F8	*Sargochromis carlottae?*	Lukanga Swamps Kafue River
CK270	2288	UZ1F9/CK281	*Sargochromis species*	Lukanga Swamps Kafue River
CK271	-	UZ1D'	*Sargochromis species*	Lukanga Swamps Kafue River
CK272	-	UZID"	*Sargochromis species*	Lukanga Swamps Kafue River
CK273	2257	UZ1C4	*Sargochromis species*	Lukanga Swamps Kafue River
CK274	2LP-277	UZ1E2/CK283	*Sargochromis species*	Lukanga Swamps Kafue River
CK275	-	UZ1C6	*Sargochromis species*	Lukanga Swamps Kafue River
CK276	-	UZ1C5	*Sargochromis species*	Lukanga Swamps Kafue River
CK277	-	UZ1DVI	*Sargochromis species*	Lukanga Swamps Kafue River
CK278	-	UZ1DV	*Sargochromis species*	Lukanga Swamps Kafue River
CK279	-	UZ1F10	*Sargochromis species*	Lukanga Swamps Kafue River
CK280	-	See CK279	*Sargochromis species*	Lukanga Swamps Kafue River
CK281	2288	See CK270	*Sargochromis species*	Lukanga Swamps Kafue River
CK282	2272-2257	UZ1C3	*Sargochromis species*	Lukanga Swamps Kafue River
CK283	2277	See CK274	*Sargochromis species*	Lukanga Swamps Kafue River

Table 1 continued

CK284	2286	UZ1E6/CK258	*Sargochromis species*	Lukanga Swamps Kafue River
CK285	-	UZ1H6	*Sargochromis species*	Mazabuka Kafue River
CK286	-	UZ1H7	*Sargochromis species*	Mazabuka Kafue River
CK287	-	UZ1H8	*Sargochromis species*	Mazabuka Kafue River
CK288	-	UZ1H3	*Sargochromis carlottae?*	Mazabuka Kafue River
CK289	-	UZ1H9	*Sargochromis carlottae?*	Mazabuka Kafue River
CK290	-	UZ1H10	*Sargochromis species*	Mazabuka Kafue River
CK291	-	UZ1Hii	*Sargochromis species*	Mazabuka Kafue River
CK292	-	UZ1Hiii	*Sargochromis species*	Mazabuka Kafue River
CK293	-	UZ1Hiv	*Sargochromis species*	Mazabuka Kafue River
CK294	-	UZ1I1	*Sargochromis species*	Mazabuka Kafue River
CK295	-	UZ1I2	*Sargochromis species*	Mazabuka Kafue River
CK296	-	UZ1Ii	*Sargochromis species*	Mazabuka Kafue River
CK297	-	UZ1Iii	*Sargochromis species*	Mazabuka Kafue River
CK298	2647	UZ3A1	*Sargochromis giardi?*	Mukakani Camp UZR
CK299	2648	UZ3A4	*Sargochromis species*	Mukakani Camp UZR
CK300	2675	UZ3A8	*Sargochromis species*	Mukakani Camp UZR
CK301	-	UZ3C5	*Sargochromis species*	Mongu Market UZR
CK302	-	UZ3C6	*Sargochromis species*	Mongu Market UZR
CK303	-	UZ3C8	*Sargochromis species*	Mongu Market UZR
CK304	-	UZ3C10	*Sargochromis species*	Mongu Market UZR
CK305	-	UZ3D1	*Sargochromis species*	Mongu Market UZR
CK306	-	UZ3D2	*Sargochromis species*	Mongu Market UZR
CK307	-	UZ3D10	*Sargochromis species*	Likundu Camp UZR
CK308	-	UZ3E6	*Sargochromis species*	Likundu Camp UZR
CK309	-	UZ3E6b	*Sargochromis species*	Likundu Camp UZR
CK310	-	UZ3E7	*Sargochromis species*	Likundu Camp UZR
CK312	-	UZ3E8	*Sargochromis species*	Likundu Camp UZR
CK313	-	UZ3E9	*Sargochromis species*	Likundu Camp UZR
CK314	-	UZ3E10	*Sargochromis species*	Likundu Camp UZR
CK315	-	UZ3F1	*Sargochromis species*	Likundu Camp UZR
CK316	-	UZ3F2	*Sargochromis species*	Likundu Camp UZR
CK317	-	UZ3F3	*Sargochromis species*	Likundu Camp UZR
CK318	-	UZ3F4-1	*Pharyngochromis species*	Likundu Camp UZR
CK319	-	UZ3F5	*Sargochromis species*	Likundu Camp UZR
CK320	-	UZ3F6	*Sargochromis codringtonii?*	Likundu Camp UZR
CK321	-	UZ3F9	*Sargochromis carlottae?*	Likundu Camp UZR
CK322	-	UZ3F10	*Sargochromis carlottae?*	Likundu Camp UZR
CK323	-	UZ3G6	*Sargochromis codringtonii?*	Imukomba Camp UZR
CK324	-	UZ3G7	*Sargochromis species*	Imukomba Camp UZR
CK325	-	UZ3G10	*Sargochromis carlottae?*	Imukomba Camp UZR
CK326	-	UZ3H7	*Sargochromis codringtonii?*	Mukakani Camp UZR
CK327	-	UZ3H10	*Pharyngochromis species*	Nalyele Camp UZR
CK328	-	UZ3I1	*Pharyngochromis species*	Nalyele Camp UZR
CK329	-	UZ1E7	*Pseudocrenilabrus philander*	Lukanga Swamps Kafue River

Table 1 continued

CK330	2284	UZ1E8	*Pseudocrenilabrus philander*	Lukanga Swamps Kafue River
CK331	-	UZ1E9	*Pseudocrenilabrus philander*	Lukanga Swamps Kafue River
CK332	-	UZ1E10	*Pseudocrenilabrus philander*	Lukanga Swamps Kafue River
CK333	2329	UZ2C5	*Pseudocrenilabrus philander*	Lake Kariba MZR
CK334	2330	UZ2C6	*Pseudocrenilabrus philander*	Lake Kariba MZR
CK335	2331	UZ2C7	*Pseudocrenilabrus philander*	Lake Kariba MZR
CK336	2332	UZ2C8	*Pseudocrenilabrus philander*	Lake Kariba MZR
CK337	2333	UZ2C9	*Pseudocrenilabrus philander*	Lake Kariba MZR
CK338	-	UZ3EOi	*Pseudocrenilabrus philander*	Kabala UZR
CK339	-	UZ3EOii	*Pseudocrenilabrus philander*	Kabala UZR
CK340	-	UZ3E10iii	*Pseudocrenilabrus philander*	Kabala UZR
CK341	2338	UZ4I5	*Haplochromis orange*	Mwatishi /Lake Mweru
CK342	2339	UZ4I6	*Haplochromis orange*	Mwatishi /Lake Mweru
CK343	2340	UZ4I7	*Haplochromis orange*	Mwatishi /Lake Mweru
CK344	2341	UZ4I8	*Haplochromis orange*	Mwatishi /Lake Mweru
CK345	2342	UZ4I9	*Haplochromis orange*	Mwatishi /Lake Mweru
CK346	2343	UZ4I10	*Haplochromis orange*	Mwatishi /Lake Mweru
CK347	-	UZ4Ii	*Haplochromis orange*	Mwatishi /Lake Mweru
CK348	-	UZ2C10	*Pharyngochromis acuticeps*	Lake Kariba
CK349	-	UZ2D1	*Pharyngochromis acuticeps*	Lake Kariba
CK350	2361	UZ2D2	*Pharyngochromis acuticeps*	Lake Kariba
CK351	2362	UZ2D3	*Pharyngochromis acuticeps*	Lake Kariba
CK352	-	UZ2D4	*Pharyngochromis acuticeps*	Lake Kariba
CK353	-	UZ2D5	*Pharyngochromis acuticeps*	Lake Kariba
CK354	-	UZ2D6	*Pharyngochromis acuticeps*	Lake Kariba
CK355	-	UZ2D7	*Pharyngochromis acuticeps*	Lake Kariba
CK356	-	UZ2E4	*Pharyngochromis acuticeps*	Lake Kariba
CK357	2363	UZ2E5	*Pharyngochromis acuticeps*	Lake Kariba
CK358	2364	UZ2E6	*Pharyngochromis acuticeps*	Lake Kariba
CK359	-	UZ2E7	*Pharyngochromis acuticeps*	Lake Kariba
CK360	-	UZ2E8	*Pharyngochromis acuticeps*	Lake Kariba
CK361	-	UZ2Ei	*Pharyngochromis acuticeps*	Lake Kariba
CK362	2548	UZ2A3	*Serranochromis macrocephalus*	Lake Kariba
CK363	-0	UZ2A4	*Serranochromis macrocephalus*	Lake Kariba
CK364	2549	UZ2A5	*Serranochromis macrocephalus*	Lake Kariba
CK365	2550	UZ2A6	*Serranochromis macrocephalus*	Lake Kariba
CK366	2551	UZ2A7	*Serranochromis macrocephalus*	Lake Kariba
CK367	2552	UZ2A8	*Serranochromis macrocephalus*	Lake Kariba
CK368	-	UZ CK368	*Serranochromis macrocephalus*	Lake Kariba
CK369	-	UZ CK369	*Serranochromis macrocephalus*	Lake Kariba
CK370	-	UZ CK370	*Serranochromis macrocephalus*	Lake Kariba
CK371	-	UZ CK371	*Serranochromis macrocephalus*	Lake Kariba
CK372	-	UZ CK372	*Serranochromis macrocephalus*	Lake Kariba
CK373	-	UZ CK373	*Serranochromis macrocephalus*	Lake Kariba
CK374	2546	UZ2F2	*Serranochromis macroceph?*	Lake Kariba

Table 1 continued

CK375	2547	UZ2F3	*Serranochromis macrocephalus*	Lake Kariba
CK376	2663	UZ2A10	*Sargochromis codringtonii*	Lake Kariba
CK377	2664	UZ2B1	*Sargochromis codringtonii*	Lake Kariba
CK378	2665	UZ2B2	*Sargochromis codringtonii*	Lake Kariba
CK379	2666	UZ2B3	*Sargochromis codringtonii*	Lake Kariba
CK380	-	UZ2B6	*Sargochromis codringtonii*	Lake Kariba
CK381	2667	UZ2B7	*Sargochromis codringtonii*	Lake Kariba
CK382	-	UZ2B8	*Sargochromis codringtonii*	Lake Kariba
CK383	2645	UZ2E10	*Sargochromis species*	Lake Kariba
CK384	2646	UZ2F1	*Sargochromis codrington?*	Lake Kariba
CK385	-	UZ CK385	*Sargochromis codrington?*	Chanyanya Kafue River
CK386	-	UZ CK386	*Sargochromis species*	Chanyanya Kafue River
CK387	-	UZ CK387	*Sargochromis species*	Chanyanya Kafue River
CK388	-	UZ CK388	*Sargochromis species*	Chanyanya Kafue River
CK389	-	UZ1F'	*Sargochromis carlottae?*	Lukanga Swamps Kafue River
CK390	2264	See CK205	*Serranochromis macrocephalus*	Lukanga Swamps Kafue River
CK391	2276	See CK206	*Serranochromis macrocephalus*	Lukanga Swamps Kafue River
CK392	-	UZ1G10	*Sargochromis species*	Nakambala Market Mazabuka Kafue River
CK393	-	UZIGX	*Sargochromis species*	Nakambala Market Mazabuka Kafue River
CK394	-	UZ CK394	*Sargochromis species*	Nakambala Market Mazabuka Kafue River
CK395	-	UZ CK395	*Sargochromis species*	Nakambala Market Mazabuka Kafue River
CK396	-	UZ CK396	*Sargochromis species*	Nakambala Market Mazabuka Kafue River
CK397	-	UZ CK397	*Sargochromis codringtonii*	Nakambala Market Mazabuka Kafue River
CK398	-	UZ CK398	*Sargochromis species*	Nakambala Market Mazabuka Kafue River
CK399	-	UZ CK399	*Sargochromis species*	Nakambala Market Mazabuka Kafue River
CK400	-	UZ CK400	*Sargochromis species*	Nakambala Market Mazabuka Kafue River
CK401	-	UZ CK401	*Sargochromis species*	Nakambala Market Mazabuka Kafue River
CK402	-	UZ CK402	*Sargochromis species*	Nakambala Market
CK403	-	UZ CK403	*Sargochromis codringtonii?*	Nakambala Market Mazabuka Kafue River
CK404	-	UZ CK404	*Sargochromis species*	Nakambala Market Mazabuka Kafue River
CK405	-	UZ CK405	*Sargochromis species*	Nakambala Market Mazabuka Kafue River
CK406	-	UZ CK406	*Sargochromis species*	Nakambala Market Mazabuka Kafue River
CK407	-	UZ CK407	*Sargochromis species*	Nakambala Market Mazabuka Kafue River
CK408	-	UZ1H4a	*Serranochromis robustus*	Nakambala Market Mazabuka Kafue River
CK409	-	UZ1H5a	*Serranochromis robustus*	Nakambala Market Mazabuka Kafue River
CK410	-	UZ CK410	*Serranochromis macrocephalus*	Nakambala Market Mazabuka Kafue River
CK411	-	UZ1G9	*Sargochromis species*	Nakambala Market Mazabuka Kafue River
CK412	-	UZ1D'''	*Sargochromis species*	Ntumba Lukanga Swamps Kafue River
CK413	-	UZ1DIV	*Sargochromis species*	Ntumba Lukanga Swamps Kafue River
CK414	-	UZ1DVII	*Sargochromis species*	Ntumba Lukanga Swamps Kafue River
CK415	-	UZ1DVIII	*Sargochromis species*	Ntumba Lukanga Swamps Kafue River
CK416	-	UZ1DIX	*Sargochromis species*	Ntumba Lukanga Swamps Kafue River
CK417	-	UZ1DX	*Sargochromis species*	Ntumba Lukanga Swamps Kafue River
CK418	2650	UZ3F8	*Sargochromis species*	Libonda Camp Upper Zambezi River
CK419	-	UZ4G8	*Serranochromis* 'sp elongate'	Mifimbo Lake Mweru

Table 1 continued

CK420	-	UZ4G9	*Serranochromis* 'sp elongate'	Mifimbo Lake Mweru
CK421	-	UZ4G10	*Serranochromis* 'sp elongate'	Mifimbo Lake Mweru
CK422	-	UZ4H1	*Serranochromis* 'sp elongate'	Mifimbo Lake Mweru
CK423	-	UZ4H2	*Serranochromis* 'sp elongate'	Mifimbo Lake Mweru
CK424	-	UZ4I2	*Serranochromis stappersi*	Mwatishi/Lake Mweru Estuary
CK425	-	UZ4I4	*Serranochromis stappersi*	Mwatishi/Lake Mweru Estuary
CK426	-	UZ3F3a	*Sargochromis species*	Likundu Camp Upper Zambezi River
CK427	-	UZ3F3b	*Sargochromis species*	Likundu Camp Upper Zambezi River
CK428	2649	UZ3F3c	*Sargochromis species*	Likundu Camp Upper Zambezi River
CK429	-	UZ3F7	*Sargochromis codringtonii?*	Libonda Camp Upper Zambezi River
CK430	-	UZ3H1	*Sargochromis carlottae?*	Imukomba Camp Upper Zambezi River
CK431	-	UZ3K6	*Sargochromis giardi?*	Ngweze Camp Upper Zambezi River
CK432	-	UZ3K7	*Sargochromis giardi?*	Ngweze Camp Upper Zambezi River
CK433	-	UZ3K10	*Sargochromis carlottae?*	Ngweze Camp Upper Zambezi River
CK434	-	UZ3M3	*Sargochromis giardi?*	Kabala Camp Upper Zambezi River
CK435	2677	UZ3N5-1	*Sargochromis giardi?*	Kabala Camp Upper Zambezi River
CK436	-	UZ3N5-2	*Sargochromis giardi?*	Kabala Camp Upper Zambezi River
CK437	-	UZ3N5-3	*Sargochromis giardi?*	Kabala Camp Upper Zambezi River
CK438	-	UZ3N5-4	*Sargochromis giardi?*	Kabala Camp Upper Zambezi River
CK439	-	UZ3N6	*Sargochromis carlottae?*	Kabala Camp Upper Zambezi River
CK440	-	UZ3F4-2	*Pharyngochromis species*	Likundu Camp Upper Zambezi River
CK441	-	UZ3G1	*Pharyngochromis species*	Mongu Market Upper Zambezi River
CK442	-	UZ3C7	*Sargochromis species*	Mongu Market Upper Zambezi River
CK443	-	UZ3C9	*Sargochromis species*	Mongu Market Upper Zambezi River
CK444	-	UZ3I6	*Sargochromis carlottae?*	Ngweze Camp Upper Zambezi River
CK445	-	UZ3I7	*Sargochromis carlottae?*	Ngweze Camp Upper Zambezi River
CK446	2652	UZ3J3	*Sargochromis carlottae?*	Ngweze Camp Upper Zambezi River
CK447	-	UZ3J4	*Sargochromis carlottae?*	Ngweze Camp Upper Zambezi River
CK448	-	UZ3J9	*Pharyngochromis species*	Ngweze Camp Upper Zambezi River
CK449	-	UZ3J10	*Pharyngochromis species*	Ngweze Camp Upper Zambezi River
CK450	-	UZ3K1	*Pharyngochromis species*	Ngweze Camp Upper Zambezi River
CK451	-	UZ3M2	*Sargochromis carlottae?*	Kabala Camp Upper Zambezi River
CK452	-	UZ3N7-1	*Sargochromis carlottae*	Kabala Camp Upper Zambezi River
CK453	-	UZ3N7-2	*Sargochromis carlottae*	Kabala Camp Upper Zambezi River
CK454	-	UZ3N8	*Sargochromis carlottae?*	Kabala Camp Upper Zambezi River
CK455	-	UZ3N9-1	*Pharyngochromis species*	Kabala Camp Upper Zambezi River
CK456	-	UZ3N9-2	*Pharyngochromis species*	Kabala Camp Upper Zambezi River
CK457	-	UZ3N10	*Sargochromis sp. 'straight'*	Kabala Camp UZR
CK458	-	UZ CK458	*Serranochromis robustus*	Nalyele Camp Upper Zambezi
CK459	-	UZ CK459	*Serranochromis robustus*	Nalyele Camp Upper Zambezi
CK460	-	UZ CK460	*Serranochromis robustus*	Likundu Camp Upper Zambezi River
CK461	-	UZ CK461	*Serranochromis robustus*	Likundu Camp Upper Zambezi River
CK462	-	UZ CK462	*Serranochromis robustus*	Likundu Camp Upper Zambezi River
CK463	-	UZ CK463	*Serranochromis macrocephalus*	Likundu Camp Upper Zambezi River
CK464	2534	UZ CK464	*Serranochromis macrocephalus*	Likundu Camp Upper Zambezi River

Table 1 continued

CK465	-	UZ3G8	*Serranochromis macrocephalus*	Mongu Market Upper Zambezi River
CK466	-	UZ3G9	*Serranochromis macrocephalus*	Mongu Market Upper Zambezi River
CK467	2678	UZ3H4	*Serranochromis altus?*	Mongu Market Upper Zambezi River
CK468	2679	UZ3H5	*Serranochromis altus*	Mongu Market Upper Zambezi River
CK469	-	UZ3G2	*Serranochromis macrocephalus*	Mongu Market Upper Zambezi River
CK470	2600	UZ3E5	*Serranochromis angusticeps*	Likundu Camp Upper Zambezi River
CK471	2676	UZ3I4	*Sargochromis giardi*	Maramba Market Livingstone Upper Zambezi
CK472	2681	UZ3I2	*Serranochromis altus?*	Maramba Market Livingstone Upper Zambezi
CK473	-	UZ3I3	*Serranochromis altus*	Maramba Market Livingstone Upper Zambezi
CK474	-	UZ3N3-1	*Serranochromis altus?*	Kabala Camp Upper Zambezi River
CK475	2686	UZ3N3-2	*Serranochromis altus?*	Kabala Camp Upper Zambezi River
CK476	-	UZ3N3-3	*Serranochromis altus?*	Kabala Camp Upper Zambezi River
CK477	2611	UZ3N4-1	*Serranochromis angusticeps*	Kabala Camp Upper Zambezi River
CK478	-	UZ3N4-2	*Serranochromis angusticeps*	Kabala Camp Upper Zambezi River
CK479	-	UZ3L10	*Serranochromis angusticeps*	Kabala Camp Upper Zambezi River
CK480	-	UZ3M1	*Serranochromis angusticeps*	Kabala Camp Upper Zambezi River
CK481	2470	UZ3K4	*Serranochromis angusticeps*	Ngweze Camp Upper Zambezi River
CK482	2610	UZ3K5	*Serranochromis angusticeps*	Ngweze Camp Upper Zambezi River
CK483	-	UZ3J2	*Serranochromis robustus*	Ngweze Camp Upper Zambezi River
CK484	2633	UZ3N1-1	*Serranochromis robustus*	Kabala Camp UZR
CK485	-	UZ3N1-2	*Serranochromis robustus*	Kabala Camp Upper Zambezi River
CK486	2631	UZ3L8	*Serranochromis robustus*	Kabala Camp Upper Zambezi River
CK487	2632	UZ3L9	*Serranochromis robustus*	Kabala Camp Upper Zambezi River
CK488	2545	UZ3I5	*Serranochromis macrocephalus*	Maramba Market Livingstone Upper Zambezi
CK489	2542	UZ3N2-1	*Serranochromis macrocephalus*	Kabala Camp Upper Zambezi River
CK490	-	UZ3N2-2	*Serranochromis macrocephalus*	Kabala Camp Upper Zambezi River
CK491	-	UZ3N2-3	*Serranochromis macrocephalus*	Kabala Camp Upper Zambezi River
CK492	-	UZ3N2-4	*Serranochromis macrocephalus*	Kabala Camp Upper Zambezi River
Ck493	-	UZ3N2-5	*Serranochromis macrocephalus*	Kabala Camp Upper Zambezi River
CK494	2540	UZ3L6	*Serranochromis macrocephalus*	Kabala Camp Upper Zambezi River
CK495	2541	UZ3L7	*Serranochromis macrocephalus*	Kabala Camp Upper Zambezi River
CK496	-	UZ CK496	*Serranochromis macrocephalus*	Kabala Camp Upper Zambezi River
CK497	-	UZ CK497	*Serranochromis macrocephalus*	Kabala Camp Upper Zambezi River
CK498	2543	UZ3I10	*Serranochromis macrocephalus*	Ngweze Camp Upper Zambezi River
CK499	2544	UZ3J1	*Serranochromis macrocephalus*	Ngweze Camp Upper Zambezi River
CK500	-	UZ3I8	*Serranochromis macrocephalus*	Ngweze Camp Upper Zambezi River
CK501	-	UZ3I9	*Serranochromis macrocephalus*	Ngweze Camp Upper Zambezi River
CK502	-	UZ3K2	*Pharyngochromis species*	Ngweze Camp Upper Zambezi River
CK503	-	UZ3K3	*Pharyngochromis species*	Ngweze Camp Upper Zambezi River
CK504	-	UZ CK504	*Pharyngochromis species*	Ngweze Camp Upper Zambezi River
CK505	-	UZ CK505	*Pharyngochromis species*	Ngweze Camp Upper Zambezi River
CK506	-	UZ CK506	*Pharyngochromis species*	Ngweze Camp Upper Zambezi River

Table 1 continued

CK507	-	UZ CK507	*Pharyngochromis species*	Ngweze Camp Upper Zambezi River
CK508	-	UZ3M7	*Pharyngochromis species*	Kabala Camp Upper Zambezi River
CK509	-	UZ3M8	*Pharyngochromis species*	Kabala Camp Upper Zambezi River
CK510	-	UZ CK510	*Pharyngochromis species*	Kabala Camp Upper Zambezi River
CK511	-	UZ CK511	*Pharyngochromis species*	Kabala Camp Upper Zambezi River
CK512	-	UZ CK512	*Pharyngochromis species*	Kabala Camp Upper Zambezi River
CK513	2639	UZ3L4	*Pharyngochromis species*	Kabala Camp Upper Zambezi River
CK514	2370	UZ3L5	*Pharyngochromis species*	Kabala Camp Upper Zambezi River
CK515	-	UZ CK515	*Pharyngochromis species*	Kabala Camp Upper Zambezi River
CK516	-	UZ CK516	*Pharyngochromis species*	Kabala Camp Upper Zambezi River
CK517	-	UZ CK517	*Pharyngochromis species*	Above Victoria Falls Upper Zambezi River
CK518	-	UZ CK518	*Pharyngochromis species*	Above Victoria Falls Upper Zambezi River
CK519	-	UZ CK519	*Pharyngochromis species*	Above Victoria Falls Upper Zambezi River
CK520	-	UZ CK520	*Pharyngochromis species*	Above Victoria Falls Upper Zambezi River
CK521	-	UZ CK521	*Pharyngochromis species*	Above Victoria Falls Upper Zambezi River
CK522	2334	UZ3L3	*Pseudocrenilabrus philander*	Kabala Camp Upper Zambezi River
CK523	-	UZ CK523	*Pseudocrenilabrus philander*	Kabala Camp Upper Zambezi River
CK524	-	UZ CK524	*Pseudocrenilabrus philander*	Kabala Camp Upper Zambezi River
CK525	-	UZ CK525	*Pseudocrenilabrus philander*	Mukula Stream Lake Bangweulu
CK526	-	UZ CK526	*Pseudocrenilabrus philander*	Mukula Stream Lake Bangweulu
CK527	-	UZ CK527	*Pseudocrenilabrus philander*	Mukula Stream Lake Bangweulu
CK528	-	UZ CK528	*Pseudocrenilabrus philander*	Mukula Stream Lake Bangweulu
CK529	-	UZ CK529	*Pseudocrenilabrus philander*	Mukula Stream Lake Bangweulu
CK530	2208	UZ-T03-1E1	*Pseudocrenilabrus philander*	Lunzua River Above Lunzua Falls
CK531	2803	UZ-T03-1E4	*Pseudocrenilabrus philander*	Lunzua River Above Lunzua Falls
CK532	2804	UZ-T03-1E5	*Pseudocrenilabrus philander*	Lunzua River Above Lunzua Falls
CK533	2809	UZ-T03-1E11	*Pseudocrenilabrus philander*	Lunzua River Above Lunzua Falls
CK534	2820	UZ-T03-1F11	*Pseudocrenilabrus philander*	Lunzua River Above Lunzua Falls
CK535	-	UZ4B9	*Serranochromis thumbergi*	Chalwe Lake Bangweulu
CK536	-	UZ4B10	*Serranochromis thumbergi*	Chalwe Lake Bangweulu
CK537	-	UZ CK537	*Serranochromis angusticeps*	Chambeshi Bridge Chambeshi River
CK538	-	UZ CK538	*Pseudocrenilabrus philander*	Chambeshi Bridge Chambeshi River
CK539	-	UZ CK539	*Pseudocrenilabrus philander*	Chambeshi Bridge Chambeshi River
CK540	-	UZ CK540	*Pseudocrenilabrus philander*	Chambeshi Bridge Chambeshi River
CK541	-	UZ CK541	*Pharyngochromis species?*	Chambeshi Bridge Chambeshi River
CK542	-	UZ CK542	*Pharyngochromis species?*	Chambeshi Bridge Chambeshi River
CK543	-	UZ-T031C3	*Pharyngochromis species?*	Chambeshi Bridge Chambeshi River
CK544	-	UZ-T031B20	*Pharyngochromis species?*	Chambeshi Bridge Chambeshi River
CK545	-	UZ-T031C1	*Pharyngochromis species?*	Chambeshi Bridge Chambeshi River
CK599	-	UZ 3F4-3	*Pharyngochromis 'deep'*	Likundu Camp Upper Zambezi River

Morph. No. is the code given to each specimen in the PCA, DNA No. is the code given to each DNA sequence derived from each corresponding specimen with (-) representing cases where no sequences were generated. (?) Represents those individual whose identities were not conclusive. The DNA sequences represented in this table are for the most variable part of the control region of the mitochondrion (Dloop).

Pseudocrenilabrus philander philander from Lukanga Swamps, Kafue River

Pseudocrenilabrus 'sp haplochromis-orange' from the Lake Mweru confluence

Pseudocrenilabrus philander 'yellow' and *Pseucrenilabrus philander* 'blue' from Lunzua River.

Figure1. Some representatives of the Genus *Pseudocrenilabrus* used in the study.

Pharyngochromis aacuticeps from Lake Kariba, Middle Zambezi River

Pharyngochromis species from Upper Zambezi River

Sargochromis giardi from Upper Zambezi River

Figure 2. Some representatives of the serranochromine cichlids used in the study.

Sargochromis carlottae from the Lukanga Swamps, Kafue River system

Sargochromis 'sp longsnout-kafue' from the Lukanga Swamps, Kafue River system

Sargochromis 'sp longsnout-zambezi' from Upper Zambezi River

Figure 3. Some representative species of the genus *Sargochromis* used in the study.

Serranochromis altus from Upper Zambezi River

Serranochromis angusticeps from Lake Mweru, Luapula River system

Serranochromis longimanus from Upper Zambezi River

Figure 4. Some representatives of the genus *Serranochromis* used in the study.

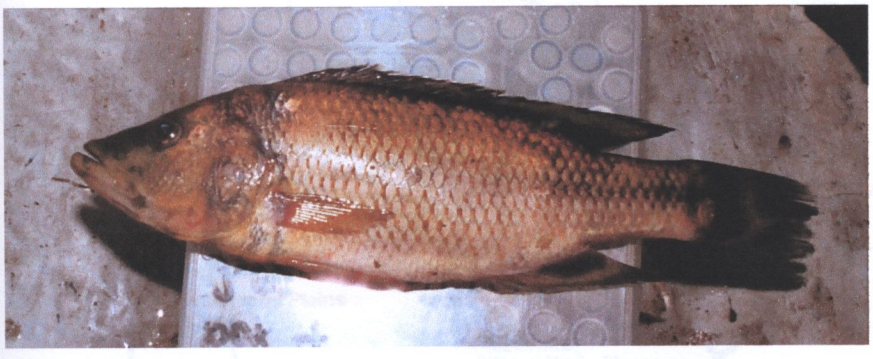
Serranochromis macrocephalus from upper Zambezi River

Serranochromis stappersi from Lake Mweru, Luapula River system

Serranochromis 'sp elongatus' a new species from Lake Mweru, Luapula River system

Figure 5. More representatives of the genus *Serranochromis* cichlids used in the study.

Serranochromis robustus from Upper Zambezi River



Serranochromis robustus? from Lake Bangweulu, Luapula River system

Serranochromis thumbergi from Lake Bangweulu, Luapula River system

Figure 6. More representative species of the genus *Serranochromis* used in the study.

Figure 7. The map of the Zambian River systems, showing the sites from where samples used in this study were collected. The names of the rivers and lakes are in black while the names of the sampling sites are shown in brown. The names of the swamps are in blue.

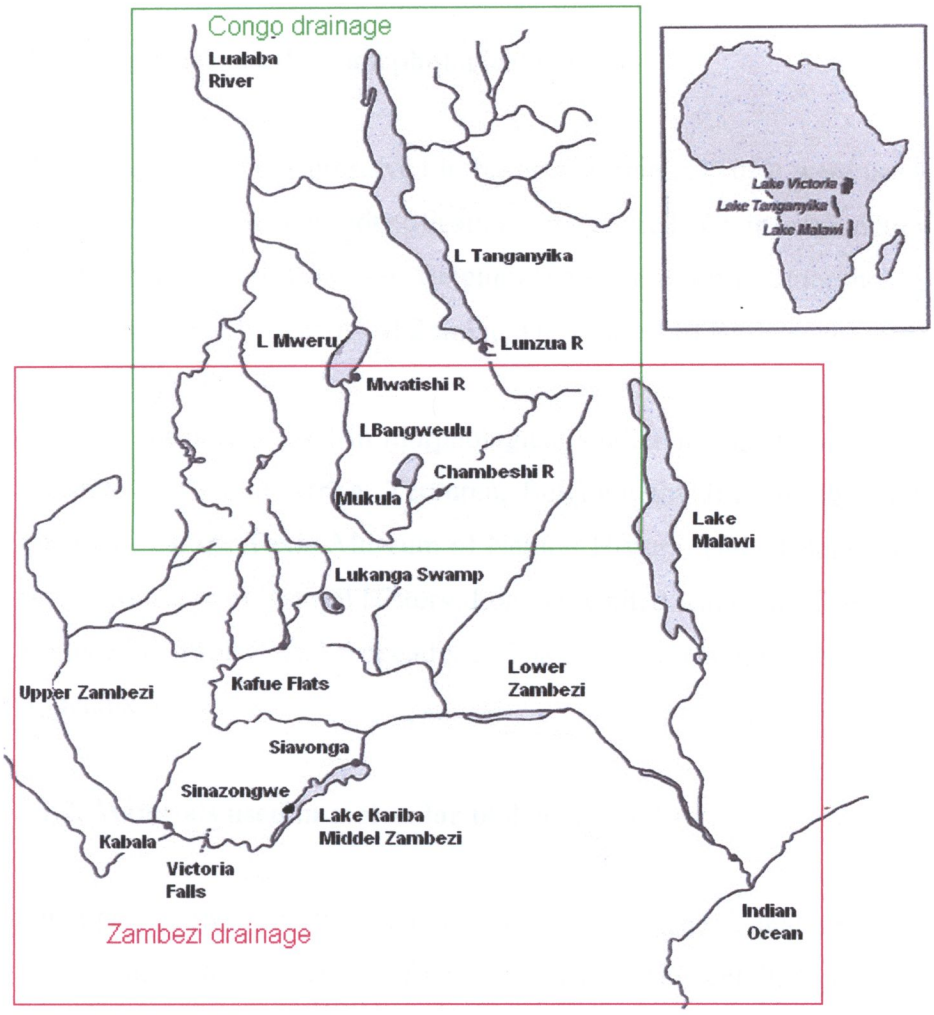


Figure 8. Sampling sites showing the relationship between the Zambezi and Congo River systems. The Congo drainage is enclosed in the green box while the Zambezi drainage is enclosed in the red box.

3.1.1. Materials used in morphological studies

Whole fish specimens preserved in buffered formalin (10%) were used for identification and description. Species identification was carried out in conjunction with the Royal African Museum at Tervuren, Belgium while comparative morphological analysis was carried out at the University of Zambia, Department of Biological Sciences

The fish specimens used in morphological studies included collections from the Royal Museum of Central Africa, Tervuren, Belgium; the JLB Smith Institute, Grahamston, South Africa; the Texas Museum of Natural History, United States of America; and the British Museum of Natural History, London, United Kingdom. The fish used in this study are listed in Table 1 and Appendix 2. The fish pictures are shown in Figures 1-6 and in Appendix 5.

3.1.2. Materials used in molecular biological studies

Fin clips preserved in 90% ethanol were used for genetic analyses. These analyses were carried out at the University of Graz in Austria. This entailed generating and analysing mitochondrial DNA sequences for all taxa in parallel to morphological characterisation. The fish samples used in molecular biological studies were exclusively from the Zambezi and Luapula Systems. DNA sequences of selected cichlids from Lake Tanganyika, obtained from the GENBANK, were used to reconstruct the Tangayika phylogeny to which the taxa used in this study belong (Salzburger *et al.*, 2002). Appendix 3 gives a list of these sequences.

3.2 Comparative morphological methods

Measurements were carried out on the fish specimens using dissecting instruments to expose the parts to be measured. A dissecting microscope was used for meristic counts while vernier calipers were used for morphometric measurements. The computer

programs MS-Excel and STATISTICA were used in recording and analyzing morphological data.

3.2.1. Morphometric data collection

Morphometric body and head measurements were performed according to Barel *et al.* (1977) with modifications according to Snoeks (1994) as in Appendix 1. See Figures 9 and 10. Measurements were started with the head using a binocular microscope followed by the other parts. A dial caliper corrected to 0.01mm was used for the measurements. Measurements were corrected up to 0.01mm for the head and up to 0.5 mm for the body. All measurements and counts were carried out from the left side of the body whenever possible. The right side was reserved for dissections and other future analyses.

3.2.2. Meristic data collection

Meristic counts were performed according to Barel *et al.* (1977) with modifications according to Snoeks (1994) as in Appendix 1. See Figure 11. Counts were done using a binocular microscope. All counts were carried out from the left side of the body whenever possible.

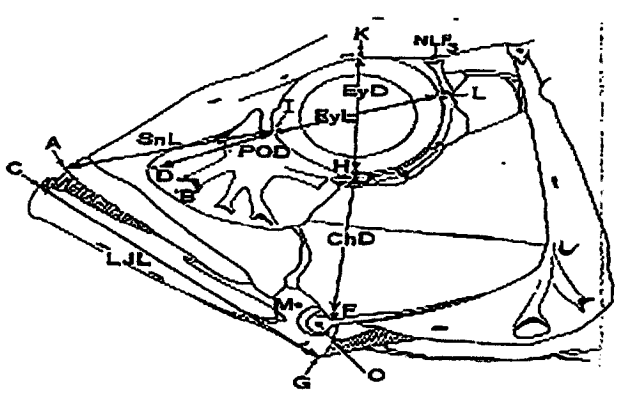


Figure 9. Determination of head measurements after Barel *et al.* (1977)
ChD = Cheek depth; EyD = Eye diameter; EyL = Eye length; POD =Pre-orbital depth; LJL = Lower jaw length; SnL = Snout length. See Appendix 1 for detailed explanations of head measurements.

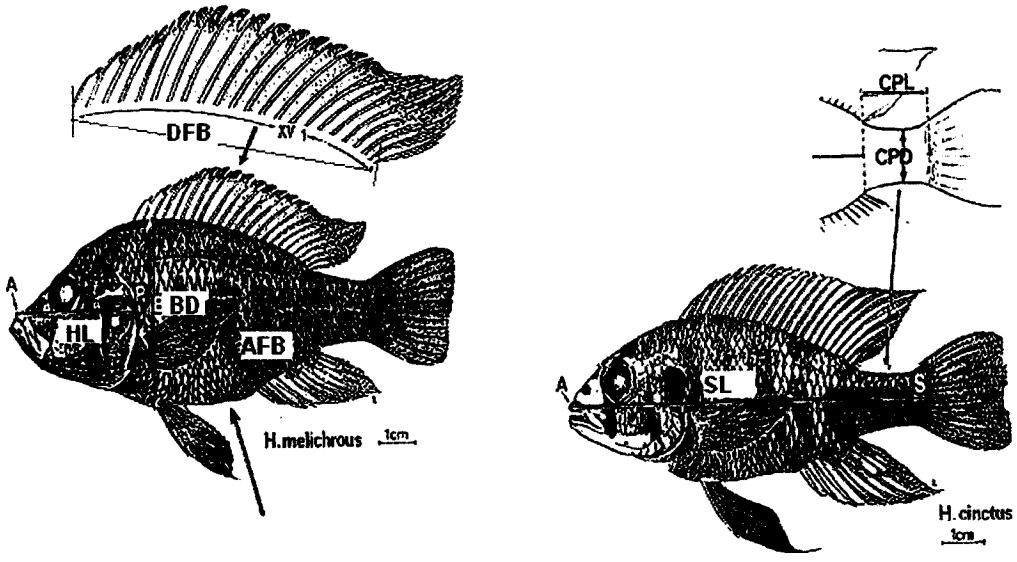


Figure 10. Determination of body measurements after Barel *et al.* (1977).

AFB = Anal fin base; DFB = Dorsal fin base; BD = Body depth; HL = Head length; SL = Standard length; CPL = Caudal peduncle length; CPD = Caudal peduncle depth. See Appendix 1 for detailed explanations of body measurements.

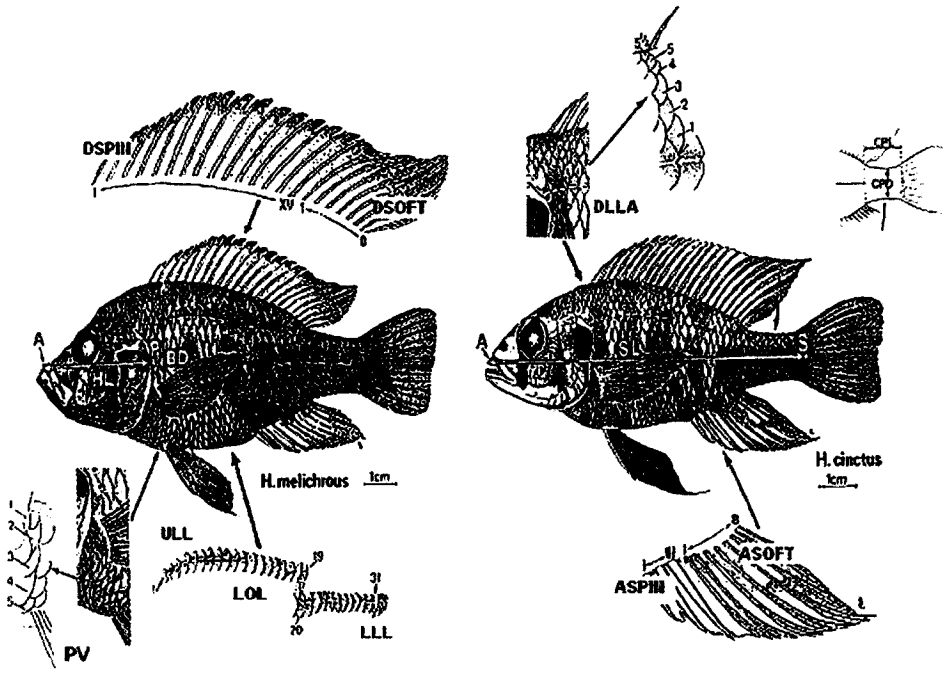


Figure 11. Determination of meristic counts after Barel *et al.* (1977). DSPIN = Dorsal fin spiny rays; DSOFT = Dorsal fin soft rays; ASPIN = Anal fin spiny rays; ASOFT = Anal fin soft rays; PV = Number of scales between the pelvic and ventral (pectoral) fins, ULL = Number of scales along the upper lateral line; LLL = Number of scales along the lower lateral line; LOL = Number of scales along the upper and lower lateral lines plus those along the diagonal between the two; DLLA = Number of scales between the dorsal fin and the upper lateral line. See Appendix 1 for detailed explanations of the merstic counts.

3.2.3. Multivariate analysis based on morphological characters

Taxonomy is descriptive and holds no hypotheses about the evolutionary relationships of the taxa under examination. The standard technique used in taxonomy is multivariate analysis (Sneath and Sokal, 1973) employing principle component analysis (PCA). PCA was originally developed to test one homogeneous group but in fish taxonomy it can be used to test more groups but only in an exploratory way (Snoeks, personal communication).

Morphological data are collected from morphometric measurements and meristic counts and these are used in principal componet analysis (PCA). Morphometric measurements are then transformed into percentages and logarithms before they are used in PCA. Logarithm-transformed data are used in the description of a new spescies while percentage data are important in that they provide supplementary information for comparisons, tables and identification keys. Percentage data are not normally distributed and have a larger error than the original measurements. Meristic data are used to supplement morphometric data. In PCA, factor loadings explain the importance of the variables (characters) while factor scores explain the importance of the specimens.

The morphological characters used in this study are of well-established taxonomic value and are according to Barel *et al.* (1997). Both Morphometeric and meristic data were collected. The morphometric data were converted into percentage data and logarithm-transformed data. Thus for each of the taxons analysed, i.e. the serranochromines and *Pseudocrenilabrus*, three analyses were carried out based on log-transformed data, percentage data and meristic data. Within both the serranochromines and *Pseudocrenilabrus*, the analyses carried out involved a series of steps in an attempt to separate all the possible species. Results are shown in Chapter 4.

3.2.4 The Mann Whitney U-test

The Mann Whitney U-test is used as a non-parametric test to compare one variable for two taxa, at a time. It can be used, for example, to compare the body depths of two taxa to see if they are statistically different or not. This test is used to confirm differences between the two taxa where there are significant overlaps in PCA. No mean or variance is considered. The Man Whitney U-test is similar to the Student t-Test. The Student t-Test is a traditional parametric test, which requires the condition of normal distribution as a prerequisite (Quicke, 1997). To circumvent the need for normality, the Man Whitney U-test is based on ranking and does not compare distribution. Size related allometry tends to give false results, therefore, this test is done on specimens of the same length class. If it is

unavoidable to compare specimens of different sizes, it is necessary to be prudent in interpretation (Snoeks, personal communication).

In this study the Mann Whitney U-test was used in cases where there was no clear separation between different taxa or related populations after PCA. The test was performed in the comparison of the *Pseudocrenilabrus* species from the Lunzua and Lubufu Rivers and also in the comparison between *Sargochromis carlottae* and undescribed *Sargochromis* specimens referred to as *Sargochromis* 'sp codringtoni big' and *Sargochromis* 'sp double striped 1'. Tables 8-13 depict the results of the Mann Whitney U-test results based on percetage and meristic data.

3.3 Molecular biological methods

The technique and the molecule to be assayed must be matched to a carefully defined problem, assessed through pilot studies and the results subjected to appropriate statistical analyses. Only then will the full power of molecular systematics be realised. New methods of analysis relate not only to the generation of phylogenetic hypotheses, but also to ecology, biogeography, behaviour, physiology, development, epidemiology, etc. (Hillis *et al.*, 1996). Molecular characters can be protein sequences, DNA sequences or e.g. RFLP restriction sites. In this study, mitochondrial DNA sequences were used. These were the D-Loop part of the control region (Lee *et al.*, 1995) the NADH 2 gene (Kocher, 1995) and the cytochrome *b* gene. The choice of the use of the mitochondrial DNA as opposed to nuclear DNA was based on the facts that it is smaller in size and easier to extract and sequence. It is fast evolving and is therefore useful in the study of closely related organisms. It is clonally and maternally inherited and is therefore, not affected by mieotic recombination (Brown *et al.*, 1979; Birky *et al.*, 1989; Ridley, 1993; Conseguera *et al.*, 2002).

The experimental procedures involved up-to-date protocols for PCR and cycle sequencing involving mitochondrial DNA. The use of an ABI 3100 automatic sequencer

allowed an efficient throughput of samples with low error rate, to generate the number of sequences and genotype analyses required.

3.3.1. DNA extraction

Total DNA was extracted from fin clips of the fish samples collected. These samples included species and populations of the genera *Serranochromis, Sargochromis, Pharyngochromis* and *Pseudocrenilabrus*. DNA was extracted by using the Chelex method (Koblmüller *et al.*, in press).

3.3.1.1. Protocol for DNA extraction by the 5% Chelex™ 100 Method

The advantage of this method is that it is simple, fast and non-toxic. Chelex is a synthetic resin, which shows a high affinity for positively charged particles. As a strong chelator it deactivates K^+ and Mg^{2+} dependent DNA-digesting enzymes and in this way it protects the DNA from degradation.

1. To prepare 5% chelex, the ideal procedure was to dissolve 25 mg of chelex in 500 microlitres of distilled water.This preparation was done for each sample in a separate 1.5 ml eppendorf tube.

2. A small piece of fin-clip was cut (using sterile instruments) and placed into its own labelled eppendorf tube.

3. The sample tubes were then incubated under slight shaking at 56°Cfor at least 5 hours in a thermomixer.

4. The samples were then heated to 94°C for 15 minutes and then cooled to room temperature on the laboratory bench. (The heating time actually used was 25 minutes to allow the thermomixer to heat from 56 to 94°C).

5. The samples were finally stored at –20°C. Before use, the extracts were centrifuged at 15,000 rpm for 5 minutes to avoid carry over of chelex particles.

3.3.2. Polymerase Chain Reaction (PCR)

PCR is a chemical method for in vitro amplification of specific DNA segments. The principle of the PCR is similar to DNA replication in the cell, where a DNA polymerase synthesises new DNA at an existing nucleic acid matrix (Clayton, 1982; Beckenbach, 1991; Singer and Berg, 1992; Alberts *et al.*, 1994). It is possible to amplify billions of copies of a specific DNA region, if the flanking regions are known. Two synthetic 20 to 30 bp long DNA oligonucleotides, one complementary to each strand of the DNA double helix and lying on opposite sides of the DNA unit to be amplified, serve as primers for the PCR and determine the ends of the final fragment (Alberts *et al.*, 1994). The amplification is catalysed by a thermally stable DNA polymerase (Taq-polymerase), obtained from the hot-vent bacterial species *Thermus aquaticus*. This enzyme adds the deoxyribonucleoside triphosphates (dNTPs)(Gassen *et al.*, 1994). The PCR cycle consists of three steps:

1) **Denaturation** – The DNA extract is heated to 94°C to separate its complementary strands.

2) **Annealing** – When the reaction mixture is abruptly cooled from 94°C to about 50°C, the two primers hybridise (anneal) at the complementary nucleotide motifs providing target specificity when the primers border the selected target gene section. Because of the large excess of the primers, parental DNA duplexes are not formed.

3) **Extension**- When the reaction mixture is heated to about 72°C, the optimum temperature for the *Taq* polymerase, the replication of the target sequence starts at the primer and the elongation proceeds in the 5' to 3' direction.

The three steps constitute a cycle and are repeated several times. Because of the thermostability of the *Taq* polymerase, it is possible to carry out PCR in a closed container. This means that no more reagents are added after the first cycle. Even after a few cycles you get DNA fragments of defined length. After each cycle the generated double stranded DNA target-molecules are denatured to single stranded DNA molecules and serve as new matrices for the complementary primer, where the next replication starts. The reaction mixture must contain a Mg^{2+} buffer, the two primers, the *Taq* polymerase, a dilution buffer and free nucleotides in large excess. The *Taq* polymerase incorporates dNTPs at a rate of about 100 nucleotides per second.

The amount of target DNA, flanked by the primers, doubles after each cycle. Ideally the target sequence is amplified 2^n fold after n cycles. In this way target product can be formed exponentially until cocktail components become limiting. Thus, about 35 – 45 cycles are performed in a typical PCR-reaction. Because of the stringency of hybridzation at the appropriate annealing temperature, PCR is highly specific. It is very important to work under sterile conditions to avoid contamination with foreign DNA because this would lead to the amplification of this foreign DNA.

Cocktail preparation

The cocktail for the PCR was prepared for all samples allowing for one or more extra samples to correct pipetting errors. This is illustrated below for a cocktail prepared for 9 samples (allowing for one extra sample to account for pipettimg error).

Table 2. PCR cocktail

Material	1 sample (microlitres)	10 samples (microlitres)
Distilled water	6.8	68
DNTPs	1.7	17
Primer1	1.7	17
Primer2	1.7	17
Mg-Buffer (20mM)	1.7	17
Enzyme diluent	1.62	16.2
Taq polymerase	0.085	0.85

Primers used

The primers that were available for use in the analysis of the various mitochondrial regions are as in the tables below:

Table 3. Commonly used mitochondrial DNA primers

Mitochondrial region	Primer 1(forward)	Primer 2(reverse)
Cytochrome b	L 14124	H 15149
Control region (D-loop part)	Thr2 or L-ProF	TDK-D
ND2	Met	Trp

Table 4. Mitochondrial DNA primer sequences actually used in this study

Name	Oligonucleotide sequence	Reference
L14724	5'-GGAAGCTTGATATGAAAAACCATCGTTC-3'	Kocher *et al.*,1989
H15149	5'-AAACTGCAGCCCCTCAGAATGATATTTGTCCTCA-3'	Kocher *et al.*,1989
Thr-2	5'- GCTTACACCAGTCTTGTAAACC-3'	Kocher *et al.*,1989
TDK-D	5'-CCTGAAGTAGGAACCAGATG-3'	Kocher *et al.*,1989
Met	5'-CATACCCCAAC?ATGTTGGT-3'	Kocher *et al.*,1995
Trp	5'-GAGATTTTCACTCCCGCTTA-3'	Kocher *et al.*,1995
L-ProF	5'-AACTCTCACCCCTAGCTCCCAAAG-3'	Koblmüller *et al. In press.*

Procedure

The cocktail was subjected to a quick-run centrifugation of up to 3000 rpm.

1. 15.3 microlitre aliquots of the cocktail were then placed into separate, labelled eppendorf tubes.

2. 1.7 microlitres of each DNA extract (template) were added to separate tubes containing the cocktail, followed by a quick run.

3. The contents of the eppendorf tubes were then loaded into capillary tubes, taking note of the order in which they were loaded.

4. The capillaries were loaded into the PCR machine and the PCR programme was run as follows:

1 cycle 94 °C for 15 seconds – Hold (*Taq* polymerase activation and first denaturation)

40 cycles: 94 °C for 0 seconds – Denaturation
52 °C for 0 seconds – Annealing
72 °C for 15 seconds – Extension

1 cycle 72 °C for 5 minutes – complete chain elongation

The PCR was run on the Rapid Cycler (Idaho Technology) and the process took about 40 minutes. The cycler temperature was suddenly dropped from 94°C to 52 °C degrees by fast cooling and then suddenly heated from 52 °C to 72 °C by fast heating.

3.3.3. Agarose minigel analysis

This is performed to examine the success in obtaining PCR products. The negatively charged DNA molecules move varying distances through the mesh in the agrose gel, depending on their length. The separation properties of agarose gels are optimal for DNA fragments between 50 and 1000 base pairs (bp). To visualise the DNA bands under UV light in the transilluminator, ethidium bromide is added to the gel. The addition of sucrose/dye loading buffer makes the DNA sample descend (sink) faster into the loading wells. A DNA ladder is applied to one of the wells as a length standard.

a) 25 ml molten 2 per cent agarose was mixed with 2 microlitres of ethidium bromide, taking care of the fact that ethidium bromide is a carcinogenic material.

b) The mixture was then poured into a mould and allowed to set (polymerise) for 20 minutes before being placed into an electrophoresis tank containing 1x TA buffer.

c) 2 microlitres of each PCR product was then mixed with the same volume of loading buffer (concentrated sucrose solution + dye).

d) The PCR product/loading buffer mixtures were then loaded into the gel.

e) A DNA ladder (standard) was also loaded into one of the wells.

f) The minigel electrophoresis was then run at 90V, 500mA, 250 Watts for 40 minutes.

g) The gel was then examined under UV light, a digital picture taken and the image saved on the computer. A printout of the picture was also produced for immediate reference.

Figure 12 shows the DNA profile on a gel with D-loop DNA samples.

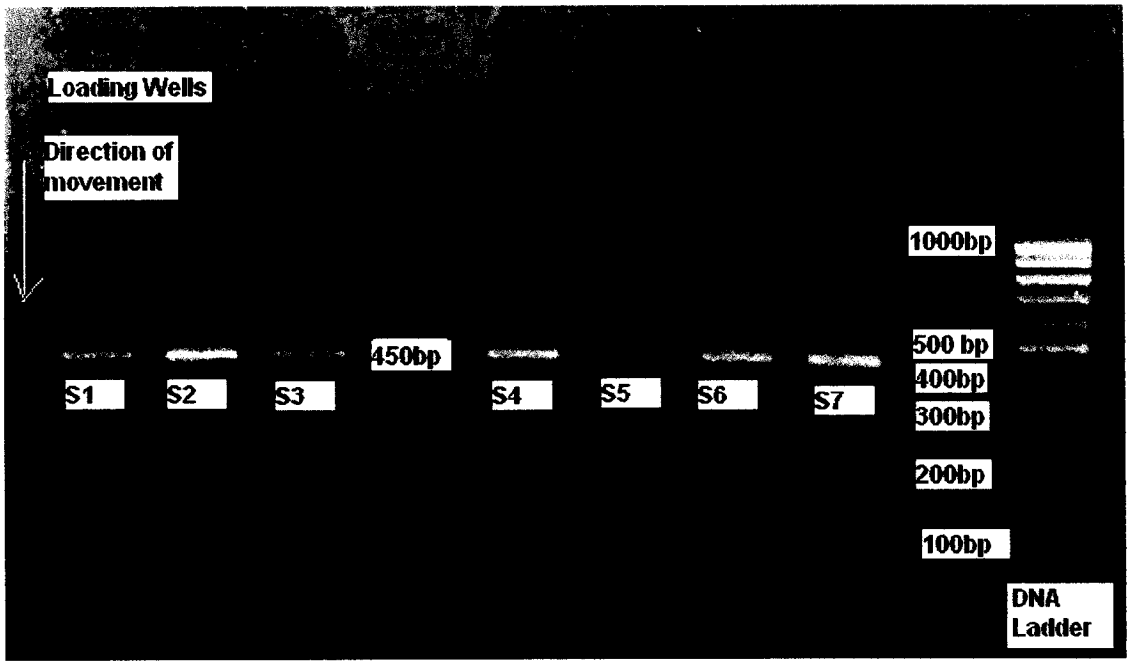


Figure 12. A minigel analysis depicting the PCR-result of the D-loop (450 bp). The DNA size-standard used is a DNA ladder with fragments ranging from 100bp to 1000bp. The samples are labeled S1-S7. The sample DNA and the ladder are loaded into the wells and allowed to move up the gel in the direction indicated.

3.3.4. Purification of PCR products using ExoSAP-IT™

This is done to remove excess primers, nucleotides, *Taq* polymerase, buffer and enzyme diluent. ExoSAP-IT™ utilises Exonuclease I and Shrimp alkaline phosphatase, to remove unconsumed dNTPs and primers. The Exonuclease I degrades single-stranded DNA molecules while the shrimp alkaline phosphatase hydrates residual dNTPs from the PCR mixture that would interfere with the sequencing reaction. The enzymes use the same buffer as used in the PCR reaction. Heating the mixture to 80°C for 15 minutes inactivates the ExoSAP-IT™.

Purification protocol

1). 5 microlitres of each PCR product was mixed with 1.5 microlitres of Exo-SAP IT. The tubes were subjected to a quick centrifuge run.

2). The samples were incubated at 37°C for 15 minutes.

3). The samples were then incubated at 80 °C for 15 minutes. The actual time used was 23 minutes to allow the thermomixer to heat from 37 to 80°C.

4). 10 microlitres of water were then added to each sample tube with a good PCR band. Less water was added to tubes with weak bands.

5). The samples were then stored at –20 °C.

3.3.5. Chain Termination Reaction (CTR) of Purified PCR Products

This reaction resembles PCR but only one primer is used, so that only single-stranded DNA fragments are synthesised (Sanger *et al.*1977). The products are amplified with linear kinetics. In addition to dNTPs the chain termination reaction mixture (TRM) also contains dideoxy-nucleotides (ddNTPs). The ddNTPs are labelled base-specifically with different fluorescent dyes. The ddNTPs act as chain-terminating inhibitors. When the ddNTPs are integrated in the strand instead of dNTPs, no further extension of the strand is possible, because no further dNTP will bind to a ddNTP. The chain termination reaction consists of 27-30 cycles resulting in a mixture of DNA fragments with different, statistically distributed lengths (Singer and Berg, 1992). Each fragment has a base-specifically labelled ddNTP at its 3' end. For confirmation and for publication, every sequence has to be amplified in both directions.

Table 5. The cocktail prepared for 7 CTR extracts (allowing for one extra)

Material	For 1 sample	For 8 samples
Termination reaction mix	1.5 microlitres	12 microlitres
Primer	0.5	4.0
Distilled water	6.0	48

Table 6. The primers that were used for the various mitochondrial regions

Mitochondrial region	Primer 1(forward)	Primer 2(backward)
Cytochrome *b*	L 14 124	H 15 149
Control region (D-loop)	Thr2 or L-ProF	TDKD
ND-2	Met	Trp

8 microlitre aliquots of the cocktail were transferred to different PCR reaction tubes. 2 microlitres of each DNA sample were added to a separate reaction tube containing the cocktail. The PCR reaction tubes were placed into the PCR machine (GeneAmp PCR System 9700 –AB) and the reaction was carried out as follows:

1 cycle of heating to hold up (H) at 94°C for 20 seconds; 28 cycles each of Denaturation (D) at 94 °C for 10 seconds, Annealing (A) at 52°C for 10 seconds and Extension (E) at 60°C for 4 seconds. Upon completion of the PCR, further incubation was allowed at 4°C until samples were transferred.

The Chain Termination Reaction products were finally transferred to labelled tubes and stored at –20°C.

3.3.6. Purification of CTR products using Sephadex suspension

Before sequencing probes can be loaded onto the automatic sequencer all unincorporated, labelled ddNTP termination molecules need to be removed. A Sephadex suspension was prepared by adding 2 mg of Sephadex powder to 32 ml distilled water. 800 microlitres of the Sephadex solution were then added to filter tubes held in filtrate chambers.

The tubes were centrifuged at 3,000 rpm for 4 minutes. Water from the lower filtrate chambers was decanted.

The pairs of tubes were assembled again and another centrifugation was done at 3,000 rpm for 1 minute. The filtrate tubes with filtrate were then put away while the filter tubes with the Sephadex column were used in the subsquent steps.

Meanwhile, 20 microlitres of distilled water were added to each sample of the CTR products and these were incubated at 56°C for five minutes.

All the contents of each sample tube were transferred onto the Sephadex column. Care was taken not to deposit the sample on tube walls and not to touch the surface of the column directly with the pipette to avoid contamination of the CTR products.

The Sephadex columns loaded with samples were placed into labelled eppendorfs and then centrifuged at 3000 rpm for 2 minutes.

The Sephadex columns were then put aside and the eppendorfs with filtrate were centrifuged for 30 minutes in a vacuum centrifuge at 45°C to give a sample DNA pellet for each eppendorf.

The pellets were stored at –20°C.

3.3.7. Automatic Sequencing

The automatic sequencing was carried out by capillary electrophoresis, performed on the 3100 genetic Analyser. The medium was a linear flowable polymer called 3100 POP-6. This polymer can be pumped out of the capillaries for each new run.

During electrophoresis in the automatic sequencing step, the shortest strands arrive first at the scanning region, where the laser is situated. The ddNTP end of each strand is bound to a dye during the CTR step. When the DNA fragments reach the detection window, the laser beam excites the dye molecules and causes them to fluoresce. The emitted base-specific fluorescence signals are detected by a photomultiplier tube and

converted into electrical signals. The electrical signals are transmitted to the computer where the bases are decoded by the sequencing software (Sequence Navigator).

The fluorescence emissions from the (16) samples are collected simultaneously and spectrally separated by a reflective spectrograph. The emissions are focused as columns of light onto the charge-coupled device (CCD) that is part of the CCD camera. The sequencing data collecting software reads and interprets the fluorescence data and then displays the data as an electropherogram.

The advantages of capillary electrophoresis on the 3100 Genetic Analyser are:
Rapid separation of DNA fragments, No gel pouring required. The 3100 genetic Analyser uses a liquid polymer, which is flowable and replenishable. There is no manual sample loading required, no gel tracking required and there is increased sensitivity resulting from the simultaneous detection of 16 samples. The method of heat dissipation for capillary electrophoresis is done evenly and efficiently because there is a lot of surface area in a capillary. This allows high voltages to be applied to migrating extension fragments without loss of resolution.

3.3.8. Sequence Alignment

The Sequence Navigator programme was used for initial comparative alignment while the adjustment was iteratively adjusted by eye, based on sequences from sister taxa, which were, themselves, aligned by Clustal W (Thompson *et. al*, 1994, Salzburger *et al.*, 2002). The aligned sequences were then analysed using PAUP 4.0b2a (Swofford, 2000).

3.3.9. Phylogenetic inference

3.3.9.1. Algorithms used in phylogenetic inference

3.3.9.2. Parsimony methods

Of the existing numerical approaches to inferring phylogenies directly from character data, methods based on the principle of maximum parsimony have been the most widely used. Most biologists are familiar with the usual notion of parsimony in science, which essentially maintains that simpler hypotheses are preferable to more complicated ones and that *ad hoc* hypotheses should be avoided whenever possible. Methods for estimating trees under the criterion of parsimony equate simplicity with the explanation of attributes shared among taxa as due to their inheritance from a common ancestor (Sober, 1989). When character conflicts occur, *ad hoc* hypotheses cannot be avoided if the observed character distributions are to be explained, and assumptions of homoplasy (convergence, parallelism, or reversal) must be invoked. In general, parsimony methods inferring phylogenies operate by selecting trees that minimize the total tree length, which reflects the number of evolutionary steps (transformations from one character state to another) required to explain a given set of data (e.g. number of base mutations). A tree that minimises the total number of steps also minimises the number of extra steps (homoplasies) needed to explain the data (Hillis *et al.*, 1996).

Parsimony methods systematically search among possible trees to identify that with the minimum number of mutational steps. Unweighted parsimony treats each type of change (e.g. for transition and transversion mutations) as equally informative. Weighted parsimony gives some types of change (usually transversions) greater importance (than transitions) in selecting the best tree. Weighted parsimony usually performs better than unweighted parsimony (Hillis *et al.*, 1996).

A common observation is that transition mutations (A to G and vice versa) occur more frequently than transversion mutations (a purine base into a pyrimidine base and vice

versa) in a given gene. For some molecules, it might even be argued that transitions occur so frequently that they quickly degenerate into "noise" and should therefore be ignored altogether. A simple method for ignoring transitions is to re-code the four nucleotides as either R (purine; A or G) or Y (pyrimidine; C or T). A disadvantage of the complete rejection of information on transitions is that, while transitions may become saturated over long evolutionary time periods, they may nonetheless be highly informative with respect to relationships among closely related taxa. One way around this dilemma is to assign greater weight to transversions than transitions without going to the extent of giving transitions zero weight (Hillis *et al.*, 1996).

All the parsimony variants can be subsumed into a generalized method that assigns a cost for the transformation of each character state to the other possible states (Sankoff, 1975). The costs can be represented as an m-by-m matrix S, where S_{ij} represents the increase in tree length (weight) associated with transformation from state i to state j, and m is the total number of possible states. An exact, dynamic programming algorithm can be used to determine the minimum length required on a given tree for any particular choice of costs and to obtain one or all of the most parsimonious reconstructions (MPRs) that yield this length (Sankoff and Cedergren, 1973; Swofford and Maddison, 1992).

3.3.9.3. Likelihood methods

3.3.9.3.1. Maximum Likelihood

This method assumes a model of nucleotide or amino acid substitution and, based on this model, identifies the tree that maximizes the probability of obtaining the observed sequences. Intuitively appealing but computationally intensive, this method is quite tolerant of violation of its assumptions and performs quite well even when substitution rates differ moderately in different branches (Yang, 1994, 1995; Olsen *et al.*, 1994).

Maximum likelihood methods of phylogenetic inference evaluate a hypothesis about evolutionary history in terms of the probability that a proposed model of the evolutionary

process and the hypothesised history will give rise to the observed data. Cavalli-Sforza and Edwards (1967) were the first to use the Maximum likelihood method in phylogenetic inference. Felsenstein (1981, 1993) brought the maximum likelihood framework to nucleotide-based phylogenetic inference. Later, maximum likelihood was applied to amino acid sequence data as well (Kishino *et al.*, 1990; Adachi and Hasegawa, 1994). In addition to its consistency properties, maximum likelihood is useful because it often yields estimates that have lower variance than other methods (i.e. it is frequently the estimation method least affected by sampling error). It also tends to be robust to many violations of the assumptions used in its models. Using this method, the major components determining the evolution of sequences can be described by just a few parameters. Even with very short sequences, maximum likelihood tree inference tends to outperform alternative methods (e.g. parsimony or additive distances) when evaluated under many models of sequence evolution (Hasegawa and Fujiwara, 1993; Kuhner and Felsenstein, 1994; Huelsenbeck, 1995). Several areas of biological research, notably genetic mapping and clinical testing, routinely use maximum likelihood methods for testing hypotheses. However, the perceived and actual complexities of obtaining maximum likelihood solutions to problems that involve numerous alternative hypotheses has inhibited the more general use of these techniques, due to computational limitations (Hillis *et al.*, 1996).

Phylogentic analysis seeks to infer the history that is most consistent with a set of observed data. In this context, the data are the base sequences; the unknowns are the branching order and branching lengths of the tree. To apply a maximum likelihood approach, a concrete model of the evolutionary process that accounts for the conversion of one sequence into another must be specified. Phylogenies are inferred by finding those trees that yield the highest likelihoods. Under the assumption that nucleotide sites evolve independently, it is possible to calculate the likelihood for each site separately, and combine the likelihoods into a total value at the end. To calculate the likelihood for a given site j, we must consider all possible scenarios by which the tip sequences could have evolved with some scenarios being more plausible than others (Swofford, 1996). For any given site (out of a total of 4), the node at the root might have possessed any of

the four possible bases: A, C, T or G. For each of these possibilities, the internal node might also have possessed any of the four nucleotides. Thus there are 4 x 4 =16 possibilities to consider. Since any one of these scenarios could have led to the nucleotide configuration at the tips of the tree, we must calculate the probability of each and sum them to obtain the total probabilities for each site j. Assuming that changes along different branches are independent, the probability of any single scenario is equal to the product of probabilities of the changes required by that scenario. The various probabilities depend on several assumptions about the process of nucleotide substitution, which defines a substitution model. The common assumption is that the evolutionary mechanisms responsible for sequence change constitute a homogeneous Markov process. The mathematical expression of a substitution model is a table of rates (substitutions per site per unit evolutionary distance) at which each nucleotide is replaced by each alternative nucleotide. For DNA sequences, these rates can be expressed as a 4 x 4 instantaneous rate matrix, in which each element represents the rate of change from one base to another base during some infinitesimal time period.

Almost all of the DNA substitution models proposed to date are special cases of the 4 x 4 matrix. It is usually assumed that the overall rate of change from one base to another and vice-versa in a given length of time is the same. Such models are said to be time-reversible. One bye product of time reversibility is that the likelihood of a tree generally does not depend on how the tree is rooted. Maximum likelihood estimation is therefore usually limited to the inference of unrooted trees, and other assumptions must be invoked to convert an unrooted tree into a rooted one.

The basic principles involved in calculating maximum likelihood are: (1) sequence alignment (2) unrooted tree for taxa whose sequences were aligned (3) tree after rooting at an internode (4) calculating the likelihood for a particular site - the sum of the probabilities of every possible reconstruction of ancestral states based on the model of base substitution (5) calculating the likelihood of the full tree - the product of the likelihoods at each site (6) the likelihood is usually evaluated by summing the log of likelihoods at each site and reported as log likelihood of the full tree (Hillis *et al.*, 1996).

If transversions occur much less frequently than transitions and the amount of divergence is high, transition differences are likely to approach or reach saturation. When this happens, transitions will contribute little phylogenetic information and will cause inflation of the variance of the evolutionary distance estimates. In such situations, it may be preferable to estimate the phylogeny using transversion data alone, minimising the impact of noisy transitions (Goldstein and Pollack, 1994). Formulas can be modified to estimate the number of transitions and transversions per site separately. Alternatively, one could recode the nucleotide states into R (A or G) and Y (C or T) and apply two-state distance correction. Alternative distances have been proposed for the K2P model (Kimura, 1980) that make separate estimates of the number of transition versus transversion substitutions and use a weighted combination of these as the estimate of evolutionary distance (Schöniger and von Haeseler, 1993; Goldstein and Pollock, 1994; Tajima and Takezaki, 1994). These methods appear to be much more reliable than the usual K2P distance (Pollock and Goldstein, 1994).

The maximum-likelihood method of reconstructing phylogenetic relationships (Felsenstein, 1981) has become more popular recently because of new powerful computers. The main advantage of this method is the application of a well-defined model of sequence evolution to a given set of data (Strimmer and von Haeseler, 1997)

3.3.9.3.2. Quartet Puzzling

This is a maximum likelihood method used, instead of bootstrapping, to test the topology of the maximum likelihood tree. Quartet puzzling is employed to reconstruct the topology (branching pattern) of a phylogenetic tree based on DNA or amino acid sequence data. This method applies maximum-likelihood tree reconstruction to all possible quartets that can be formed from n sequences. The quartet trees serve as starting points to reconstruct a set of optimal n-taxon trees. The majority rule consensus of these trees defines the quartet-puzzling tree and shows groupings that are well supported. The performance of quartet puzzling to construct trees is always equal to or better than that of neighbour joining. For cases with high transition/transversion bias, quartet puzzling outperforms

neighbour joining by a factor of 10. A PHYLIP-compatible ANSIC program, PUZZLE for analysing nucleotide or amino acid sequence data is available (Strimmer and von Haeseler, 1997).

In maximum-likelihood analysis as in maximum parsimony, the number of possible tree topologies grows exponentially with the number of sequences. Therefore, all methods have to rely on heuristic searches to find the best topology. This method reconstructs the maximum-likelihood tree for each of the (*n, 4*) possible quartets. In the puzzling step, the resulting quartet trees are combined to an overall tree. Sequences are added sequentially at random to an already existing sub-tree. The position of a new sequence is determined by a voting procedure, considering all quartets. Finally, an intermediate tree relating *n* sequences is obtained. In general, there is no *n*-taxon tree that fits all the (*n, 4*) different quartet trees. Therefore, the puzzling step is repeated several times, thereby elucidating landscape of possible optimal trees. The quartet-puzzling tree is obtained as a majority-rule consensus (Margush and Mc Morris, 1981) of all trees that result from multiple runs of the puzzling step. This tree may be binary or multifurcating. In addition to the tree topology, the quartet-puzzling tree shows reliability values for each internal branch. The algorithm is as follows: Quartet puzzling is essentially a three-step procedure, first reconstructing all possible quartet maximum likelihood trees (maximum-likelihood step), then repeatedly combining the quartet trees to an overall tree (puzzling step) and finally computing the majority rule consensus of all intermediate trees giving the quartet puzzling tree (consensus step).

Quartet puzzling, therefore, is a simple method to get a phylogenetic tree and simultaneously an impression of how well the data are suited for a phylogenetic reconstruction (Margush and Mc Morris, 1981; Strimmer and von Haeseler, 1996).

3.3.9.3.3. Four Cluster Likelihood Mapping Analysis (4CLMA)

This is a graphical method of visualizing the phylogenetic content of a set of aligned sequences. The method of analysis of the maximum likelihoods for the three fully

resolved tree topologies that can be computed for four base sequences is called Maximum Likelihood Mapping. The three likelihoods are represented as one point inside an equilateral triangle. The triangle is partitioned in different regions. One region represents star-like evolution (no phylogenetic signal), three regions represent a well-resolved phylogeny, and three regions reflect the situation where it is difficult to distinguish between two of the three trees. The location of the likelihood values in the triangle defines the strength of the phylogenetic signal in the data set. If n sequences are analysed, then the likelihoods for each subset of four sequences are mapped into the triangle. The resulting distribution of points shows whether the data are suitable for a phylogenetic reconstruction or not (Strimmer and von Haeseler, 1997). This method helps determine whether the data generate a phylogenetic signal or not.

Likelihood mapping can also be used as a complementary approach to existing methods of *a posteriori* evaluation of tree-likeness. This method may be helpful when analysing controversial phylogenies. Similar to statistical geometry in sequence space, likelihood mapping is based on the analysis of quartets, the basic ingredients to reconstruct trees. Moreover, the description of seven basins of attraction that can be characterised as fully resolved or intermediate between two trees is also of great importance in the quartet-puzzling tree search algorithm. Using a variant of likelihood mapping it is also possible to detect recombination (Strimmer and von Haeseler, 1997).

3.3.9.4. Distance methods

These methods are based on the pairwise differences between the sequences, corrected for multiple hits. They include:

3.3.9.4.1. Unweighted pair-group method with arithmetic mean (UPGMA).

This is a very widely used method because it assumes a constant rate of evolution in each branch but performs poorly when this assumption (constant rate of molecular evolution) is violated, as it often is (Hartl, 2000).

3.3.9.4.2. The Neighbour-Joining method

Neighbour joining (Saitou and Nei, 1987) is conceptually related to traditional cluster analysis, but removes the assumption that the data are ultrametric. It does not assume that all lineages have diverged equal amounts. It, however, assumes that the data come close to fitting an additive tree, so correction for superimposed substitutions is important for data that might include lineage-to-lineage differences in average rate (Hillis *et al.*, 1996). This method essentially groups the most closely related pairs of sequences. This method is extremely efficient computationally and usually yields trees close to the minimum evolution tree (Hartl, 2000).

The neighbour joining tree keeps track of nodes on a tree rather than taxa or clusters of taxa. The raw data are provided as a distance matrix, and the initial tree is a star tree. A modified distance matrix is constructed in which the separation between each pair of nodes is adjusted on the basis of their average divergence from all other nodes. Conceptually, this adjustment has the effect of normalising the divergence of each taxon for its average clock rate. The tree is constructed by linking the least-distant pair of nodes as defined by this modified matrix. When two nodes are linked, their common ancestral node is added to the tree and the terminal nodes with respective branches are removed from the tree. This pruning process converts the newly added common ancestor into a terminal node on a tree of reduced size. At each stage in the process, a new one, corresponding to an internal node on the final tree, replaces two terminal nodes. The process is completed when two nodes remain, separated by a single branch.

A neighbour-joining algorithm seeks to represent the data by an additive tree. It can assign a negative length to a branch. Kuhner and Felsenstein (1994) modified the algorithm so that when a negative branch occurred, it was set to zero, and the difference was transferred to the adjacent branch length so the total distance between an adjacent pair of terminal nodes is unaffected. This change does not alter the topology of the tree found by the algorithm; it just guarantees non-negativity of branch lengths as estimated numbers of substitutions.

Neighbour joining is classified as an algorithmic method because it constructs only one tree and does not explicitly optimise any objective function (the branch-length estimates from neighbour joining are not, in general, optimal for the minimum evolution criterion). It should be thought of as a means of getting a starting tree for more thorough searches using branch swapping under the minimum evolution or other additive-tree criteria, not as a method for choosing a final tree.

3.3.9.4.3. Minimum evolution

Minimum evolution is the exact algorithm of the distance method implemented in Neighbour Joining. This technique examines every possible tree and selects the one that minimizes the total branch lengths. This approach is computationally intractable for a large number of sequences because there are so many possible, unrooted bifurcating trees possible (Hartl, 2000).

3.3.9.5. Choosing the right model for the construction of the phylogenetic tree

In a phylogenetic analysis, the model selection and evaluation are interrelated. There are two main criteria for evaluating a phylogenetic model: how well it fits the data at hand, and how well it fits with other reliable data (sometimes called congruence in the case of comparing trees). In selecting a model based on fit of data at hand, there are tradeoffs to consider. It is always possible to improve the apparent fit of a model by adding additional parameters. This also leads to higher sampling variances. Measures of fit are useful in deciding whether it is worth adding an extra parameter (Miller, 1990). The general approach is to choose a goodness-of-fit statistic and then search for a model that maximizes this statistic without adding unnecessary parameters that do little more than explain random fluctuations in the data. If we can assume that sites in the sequence evolve independently, then the data represent a multinomial sample, so goodness-of-fit statistics such as chi squared or the log likelihood ratio test (Sokal and Rohlf, 1981) can be used to measure the fit of the observed to the predictions of the model. In

phylogenetics it is more common to use the likelihood ratio statistic, which (unlike the chi-squared statistic) does not require the expected probability of all distinct nucleotide patterns to be calculated. A related measure, the Akaike information criterion (Akaike, 1974), can also be used to choose the most appropriete model (Posada and Crandall, 1998).

Phylogenetic inferences are premised on the inheritance of ancestral characteristics, and on the existence of an evolutionary history defined by changes in these characteristics. Although almost all methods accept the appropriateness of a tree-like model of evolution, many commonly used methods of phylogenetic inference are not explicitly based on a set of evolutionary assumptions (Hillis *et al.*, 1996).

3.3.9.6. Assessing the reliability of individual branches in a phylogenetic tree

It is important to assess the reliability of the individual internal branches of an estimated tree. Many methods have been suggested, several of which have been proposed for testing whether a particular internal branch length is significantly greater than zero in an additive distance (Li and Gouy, 1991). Non-parametric approaches, such as bootstrapping, have been widely used for testing the degree of support for particular branches.

3.3.9.6.1. Bootstrap analysis

Boostrapping allows us to generate a series of pseudosamples (by resampling the unique data set with replacement), which can be used in place of the actual samples to estimate sampling variance. Pseudosamples are analysed individually, and the proportion (P) of the pseudosamples must be generated to obtain a precise estimate of P. The sampling variance follows a binomial distribution, such that $\sigma^2 = P(1-P)/n$, where n is the number of pseudosamples (Hedges, 1992). Felsenstein (1985 and 1988) originally suggested that P could be used as a measure of repeatability, or the probability that a specified internal branch would be found in an analysis of a new, independent sample of characters

(assuming that we could replay the evolutionary tape). More recently, Felsenstein and Kishino (1993) have suggested that P could be interpreted as a measure of accuracy, or the probability that the specified branch is contained in the true tree (assuming that the phylogenetic method is consistent).

Hillis and Bull (1993) examined these two interpretations of bootstrap proportions, using both simulated and known experimental phylogenies. They found that bootstrap proportions provide relatively unbiased, but highly imprecise, estimates of repeatability. They also found that bootstrap proportions provide biased estimates of accuracy. When the phylogenetic method is consistent, bootstrapping gives underestimates of accuracy at high bootstrap values. The extent of bias depends on the number of taxa, the number of characters, and the location of the internal branch in the tree (Hillis and Bull, 1993).

Bootstrapping can be used either with methods that operate on characters directly or with methods that in which character data are first transformed into distances.

Bootstrap replicates should be evaluated under an optimality criterion rather than just a tree-building algorithm. Otherwise any bias of the algorithm will artificially inflate the bootstrap proportions.

3.3.9.7. The various steps taken in the phylogenetic inference in this study

In this study the DNA sequences of two mitochodrial genes (ND2 with 1, 047bp and cytochrome with 402 bp) and the most variable part of the mitochondrial control region (D-Loop with 368bp) from representatives of the Zambian *Pseudocrenilabrus* species and the serranochromine cichlids (riverine haplochromines) were analysed.

In addition to the sequences obtained from the riverine haplochromines (Table 1), sequences from Lake Tanganyika cichlids (Salzburger *et al.*, 2002 and Appendix 3) were included in the first analysis to test the monophyly of the riverine cichlids and to relate them to the H-lineage within the Tanganyikan cichlid phylogeny (as defined by Nishida, 1991). The Tanganyika phylogeny (with the target taxa of this study included) was then

used to select the outgroups for further analyses of the *Pseudocrenilabrus* and the serranochromines. Figures 68 and 69 depict the Tanganika phylogeny showing the placement of the *Pseudocrenilabrus* and serranochromines. The results of these analyses are presented in Chapter 4.

For both the *Pseudocrenilabrus* and serranochromines, Maximum Parsimony (MP), Neighbour Joining (NJ), and Maximum Likelihood (ML) analyses were carried out in PAUP 4.0 (Swofford and Maddison, 1992) after selecting the appropriate models of base substitution (or weighting scales) for the data set. The reliability of the NJ and the MP trees was tested by using the boostrap analysis performed with PAUP 4.0 (Swofford and Maddison, 1992). The reliability of the ML trees was tested by using quartet-puzzling analysis performed with PUZZLE 4.0 (Strimmer and von Haeseler, 1997). This was done because carrying out bootstrap analysis on the ML tree demands much more computing time than the quartet puzzling analyis and yet the latter method gives reliable results.

For MP analysis, the Sliding Window analysis (Sturmbauer *et al.*, 1996) was performed in order to define the various regions of the D-loop. Three regions were identified and defined as having; low variability (LowVar with <10% substitution rate), high variability (HighVar with 10-20% substitution rate) and hyper high variability (HyperVar with >20% substitution rate) and these were defined accordingly in the analysis. Insertions and deletions (indels) were considered as the fifth base for the D-Loop data. The ND2 data set had no indels. Transition (TS) and transversion (TV) mutations were weighted differently, with TV mutations given a higher weighting. The TS/TV ratio for each data set was calculated based on pairwise distance data generated in PAUP 4.0 (Swofford and Maddison, 1992) and subjected to EXCEL analysis. The TS/TV ratios used for the various analyses are in Table 7. 1, 000 replicates were used in the heuristic search executed using PAUP 4.0 (Swofford and Maddison, 1992). The various MP trees generated are included in Chapter 4.

Table 7. Transition/Transversion ratios determined for the *Pseudocrenilabrus* and serranochromines data used in this study.

DNA sequence	Taxon	Region or position		
D-Loop		LowVar	HighVar	HyperVar
	Pseudocrenilabrus	3: 2	2: 1	-
	Serranochromines	4: 1	4: 1	3: 1
		1st and 2nd positions	**3rd position**	
ND2	*Pseudocrenilabrus*	1: 1	7: 1	-
	Serranochromines	1: 1	9: 1	-

In the case of the NJ and ML analyses, an appropriate model was then chosen after subjecting each of the *Pseudocrenilabrus* and serranochromine data set to the MODELTEST program (Posada and Crandall, 1998), which is able to select the most appropriate model after testing the goodness-of-fit for the various models. The appropriate model is indicated for each of the NJ and ML phylogenies in Chapter 4 accordingly.

The amount of phylogenetic content in each of the data sets was tested by using the 4-cluster likelihood mapping performed with PUZZLE 4.0 (Strimmer and von Haeseler, 1997).

The linearized trees were reconstructed based on the D-Loop data of the *Pseudocrenilabrus* and serranochromines with each haplotype represented only once. This was done in order to determine the sequences of the major cladogenic events in the two taxa. The two-cluster test implemented in the program LINTRE (Takezaki *et al.*, 1995) was performed. For each taxon, rate constancy was tested for all internal nodes in the topology using the branch length test of LINTRE and the appropriate evolutionary model as determined for the ML tree. No rate heterogeneity was detected at a high significance level ($p<0.01$) and as such all the taxa were retained for further analyses.

Since no accurate geology-based dating is available for African cichlids, this study has avoided the use of absolute age estimates and only uses the observed mean sequence divergences (as percentages). The pairwise calculations of mean sequence divergence were based on the Kimura –2-parameter model (K2P; Kimura, 1980) as used by Sturmbauer *et al.* (2001). The two linearized trees are presented in Chapter 4.

Chapter 4 Results and Discussion

4.0. Results layout

The results are presented in three parts, with the first part containing morphological data. The second part contains molecular biological data and the third part contains a combination of both the morphological and molecular biological data.

4.1. Morphological results

These results were mainly used in the identification and taxonomy of the specimens. The first set consists of data from morphometric measurements and the second part consists of data from meristic counts.

4.1.1. Morphological results on *Pseudocrenilabrus*

The initial Principal component analysis (PCA) was carried out on 47 individuals of *Pseudocrenilabrus* from eight localities of Zambia, on each of the log-transformed data (Figure 13), percentage data (Figure 14) and meristic data (Figure 15). The population from the Mwatishi River/Lake Mweru confluence completely separated from the rest in all the three PCA's. The population from Lufubu River separated out from the remaining individuals in the log-transformed PCA (with an overlap) and in the meristic PCA (completely). The four variables with the highest factor loadings on the log-transformed principal component 2 (PC2) were inter-orbital width (IOW), anal fin base (AFB), caudal peduncle depth (CPD) and caudal peduncle length (CPL). The four log-tansformed variables with the highest loading on PC3 were lachrymal depth (LacrD), anal fin base (AFB), dorsal fin base (DFB) and caudal peduncle length (CPL) (Figure 13). The five percentage-transformed variables with the highest loadings on PC1 were the ratios; body depth to standard length (BD-SL), eye diameter to head length (ED-HL), caudal peduncle depth to standard length (CPD-SL), caudal peduncle depth to caudal peduncle length (CPD-CPL) and inter-orbital width to head width (IOW-HW). The four percentage-

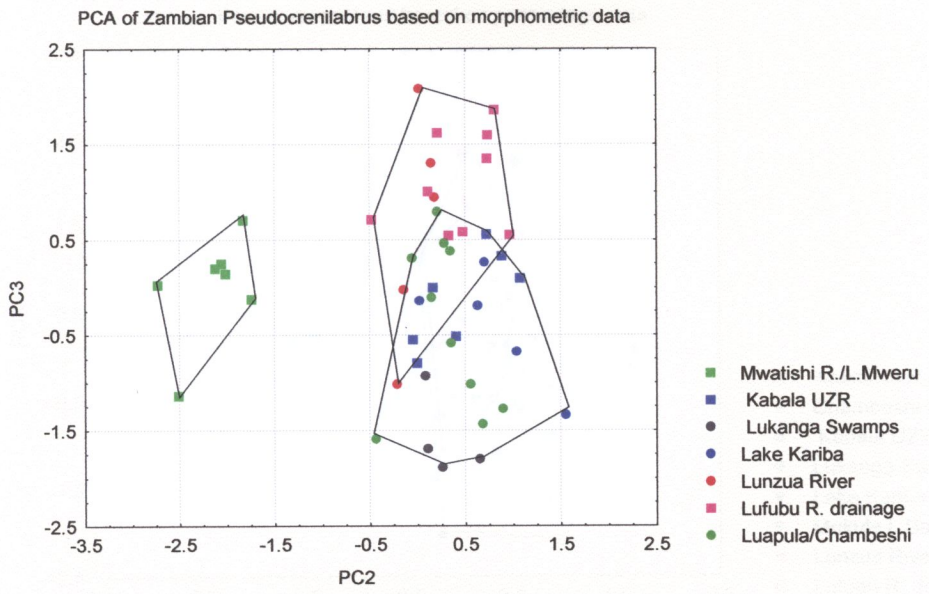
transformed variables with the highest loadings on PC2 were the ratios; pre-pelvic length to standard length (PrP-SL), dorsal fin base to standard length (DFB-SL), head length to standard length (HL-SL) and pre-ventral length to standard length (PrV-SL). The four meristic variables with the highest loadings on PC 1 were: dorsal to lateral line scales (DLLA), number of ceratobranchial gill rakers (Cerato), number of epibranchial gill rakers (EPI) and the number of scales around the caudal peduncle (CP).

The next PCA's were carried out on 17 individuals from the Lufubu and Lunzua Rivers, which is part of the Lake Tanganyika drainage. In the log-transformed PCA, there was very little separation on PC2 (Figure 16). The variables with the highest loadings on PC2 were LacrD, AFB, DFB, PrP and CDP. Figure 17 depicts the reulst of the PCA based on percentage data in which there was some reasonable separation with minor overlap on PC2. The variables with the highest loading on PC2 were CHD-HL, ED-HL, BD-SL, PrD-SL and CPL-SL. The PCA on meristic data resulted into a clear separation of the two populations on PC1. The variables with the highest loadings on PC1 were CERATO, EPI, ULL, DLLA and PV. Because of the major overlaps in the PCA's based on log-transformed and percentage data, the Mann Whitney U-test was carried out based on the percentage and meristic data and results of this test are depicted in Tables 8 and 9. Some of the variables with the highest loadings in the PCA based on percentage data are same as the variables with the highest p-value in results of the Mann Whitney test on percentage data. The results from the two methods on meristic data do not show such a trend but instead different variables are important in the two cases and as such these should then be taken in total. This means that the variables which were contributing the most to the differences between the two populations, based on meristic data, are a combination of the variables with the highest loadings in PCA and those with the highest p-value in the Mann Whitney U-test.

Then, PCA's were carried out on 22 individuals from the Zambezi and Luapula drainages using the same approach as above. The results are depicted in Figures 19-21. The variables with the highest factor loadings are also depicted in Figures 19-21. See Appendix 1 for details of the morphological characters used.

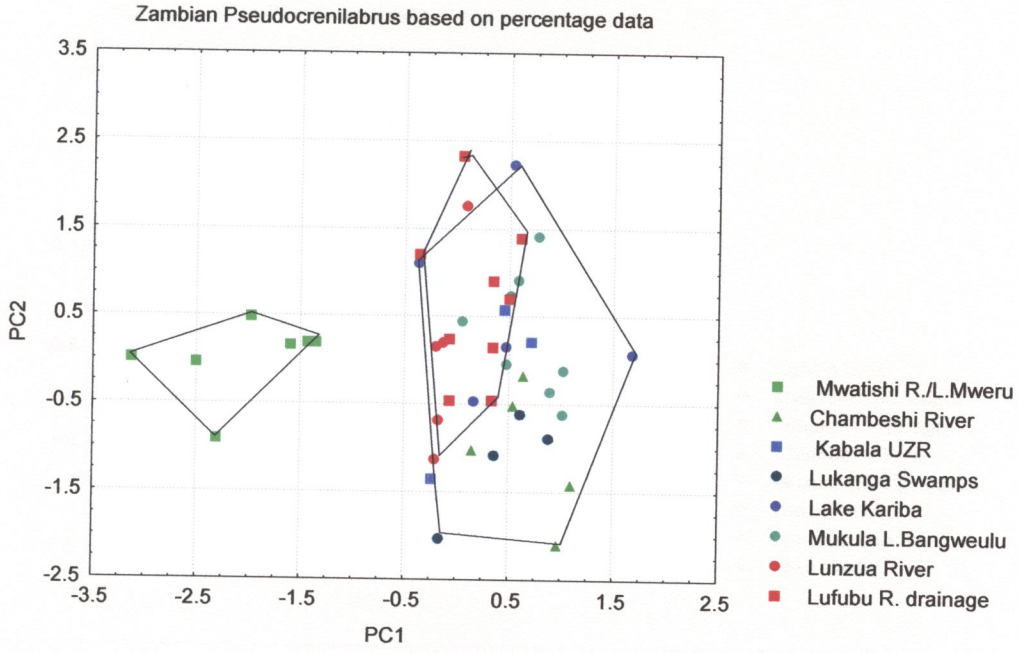
The last PCA's were carried out on 12 individuals from the Zambezi and Kafue Rivers (Figures 22-24). The variables with the highest factor loadings are depicted in figures 22-24. See also Appendix 1 for details of the characters used.

It should be noted that in all the log-transformed PCAs, the first principal component (PC1) was not included because this component is highly allometric (size-dependent) and therefore does not contain any taxonomic information. Thus only PC2 and PC3 were used. For meristic PCAs, the tooth counts are left out because the number of teeth tends to be age-related and since the age structure of the study specimens was random, the results from the number of teeth could be relied upon to provide valid taxonomic information. This is true for both *Pseudocrenilabrus* and the serranochromines.



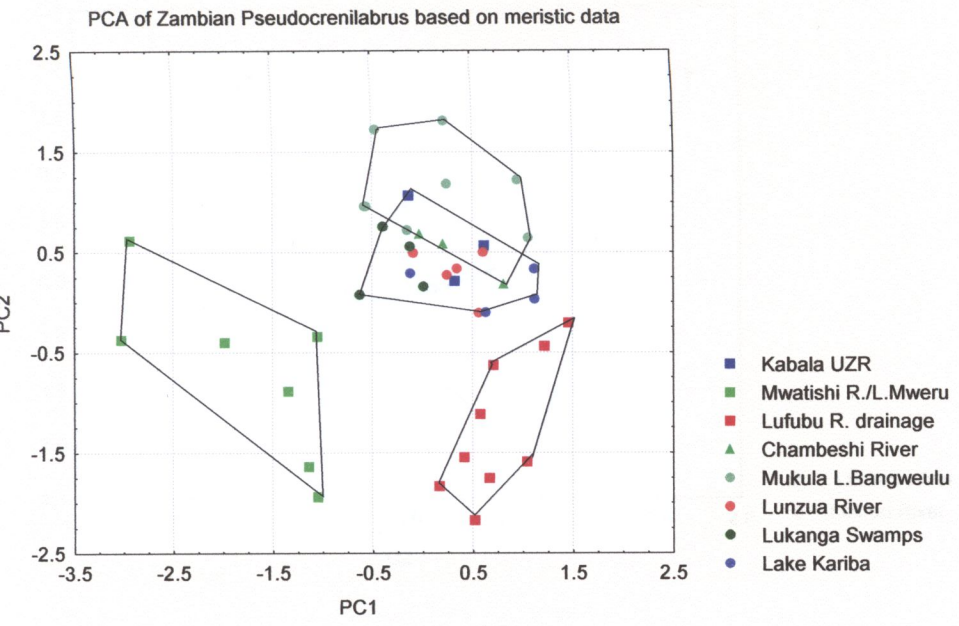
Variable	PC1	PC2	PC3	PC4	PC5
LOGLACRD	0.93188	0.082245	**0.226733**	-0.05877	0.119077
LOGSNL	0.947765	-0.02601	0.115549	-0.22794	0.027039
LOGLJL	0.949797	-0.06607	-0.0003	0.131591	-0.20839
LOGPPL	0.949027	0.061636	**0.135744**	0.04706	-0.10382
LOGCHD	0.955607	0.102302	-0.03518	-0.046	-0.17538
LOGED	0.834642	-0.4873	-0.03581	0.002253	0.033439
LOGIOW	0.846367	**0.458831**	-0.02848	0.203507	-0.01851
LOGHW	0.967144	0.079099	-0.08403	-0.05742	-0.0184
LOGHL	0.982013	0.035477	0.060293	-0.04958	-0.09137
LOGSL	0.985021	-0.10729	-0.03421	0.007831	0.056697
LOGBD	0.965023	0.165462	-0.0535	0.039297	0.041685
LOGDFB	0.961823	-0.08897	**-0.14608**	0.05921	0.097931
LOGAFB	0.875583	**-0.20993**	**-0.3903**	-0.00767	-0.05031
LOGPRD	0.967481	-0.01175	0.024223	-0.04513	0.110068
LOGPRP	0.963211	0.043034	0.070915	-0.1251	-0.05669
LOGPRV	0.980942	-0.01096	0.056442	-0.03723	-0.06412
LOGPRA	0.978248	0.004528	0.007124	-0.09017	0.04311
LOGCPL	0.860031	**-0.32091**	**0.237471**	0.264675	0.065626
LOGCPD	0.910753	**0.251634**	-0.1432	0.044568	0.207443
Proportion of total variation	0.88109	0.038994	0.018983	0.012034	0.010302

Figure 13. PCA of the Zambian *Pseudocrenilabrus* species complex. The population from the Mwatishi River/Lake Mweru confluence separates out on PC2. This is a PC2 x PC3 plot, based on log-transformed data. The number of individuals used in the analysis is 47 (N = 47). The populations from the Lufubu River drainage and Lunzua River partly separate from the populations from the Luapula-Chambeshi and the Zambezi-Kafue drainages on PC3. The variables with the highest factor loadings are set in bold and the contribution of each PC to the total variation is given in the last row.



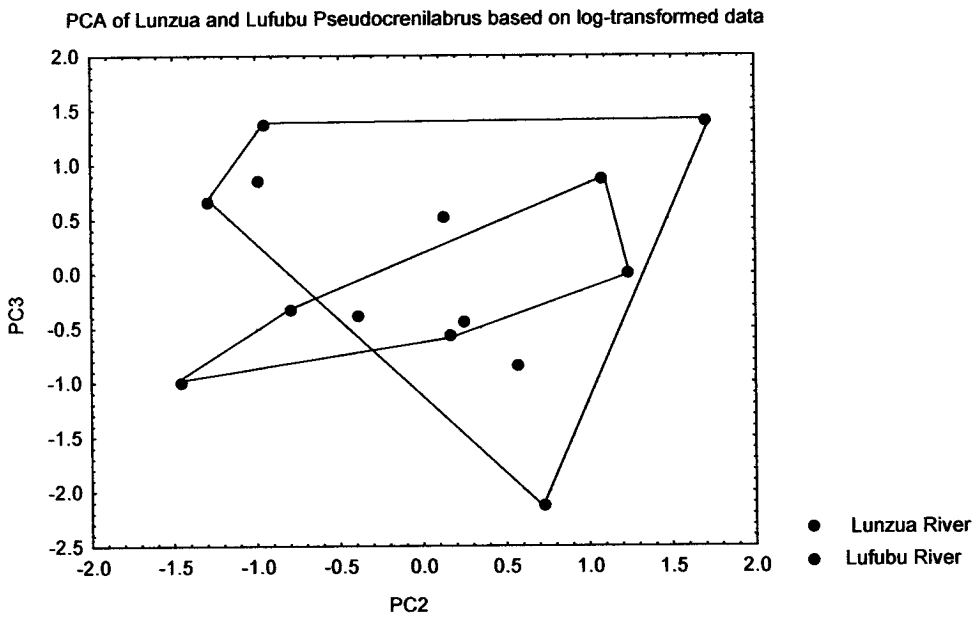
Variable	PC1	PC2	PC3	PC4	PC5
LACRD-HL	0.207775	-0.35378	-0.52608	-0.5275	0.08367
SNL-HL	-0.02276	-0.31327	-0.25858	-0.65888	-0.45269
LJL-HL	-0.14392	**-0.52175**	-0.23432	0.275608	0.021352
PPL-HL	0.201339	-0.43243	-0.64267	0.017122	-0.24816
CHD-HL	0.307681	**-0.56585**	-0.16055	-0.10047	-0.11851
ED-HL	**-0.7875**	-0.03106	0.179442	0.012156	-0.016
IOW-HL	**0.710539**	-0.39643	-0.13861	0.182101	0.454404
IOW-HW	**0.72146**	-0.12045	-0.34177	0.337311	0.321378
HW-HL	0.021384	**-0.54735**	0.41376	-0.30474	0.303881
HL-SL	0.575916	**0.632077**	-0.09201	0.223994	-0.2582
BD-SL	**0.81891**	-0.10243	-0.0678	0.2251	-0.13937
DFB-SL	0.173511	**-0.63687**	0.070674	0.348555	-0.37122
AFB-SL	-0.16693	-0.3934	0.503517	0.416293	-0.36309
PRD-SL	0.398339	0.176758	-0.11928	-0.09228	-0.40411
PRP-SL	0.478662	**0.677473**	0.080916	-0.15689	-0.03333
PRV-SL	0.46797	**0.622755**	-0.1405	0.189971	-0.23019
PRA-SL	0.539814	0.43135	0.22101	-0.44914	0.128765
CPL-SL	-0.55415	0.325551	-0.56648	0.217544	0.258468
CPD-SL	**0.762672**	-0.19305	0.228171	-0.00626	0.202754
CPD-CPL	**0.760654**	-0.28708	0.491165	-0.14265	-0.05475
Proportion of total variation	0.263569	0.186753	0.107596	0.088649	0.069551

Figure 14. PCA of the Zambian *Pseudocrenilabrus* species complex. This is a PC1 x PC2 plot, N = 47, based on percentage data. The population from the Mwatishi River/Lake Mweru confluence separates out on PC1. The Lunzua and Lufubu populations are similar to each other. The variables with the highest factor loadings are set in bold and the contribution of each PC to the total variation is given in the last row.



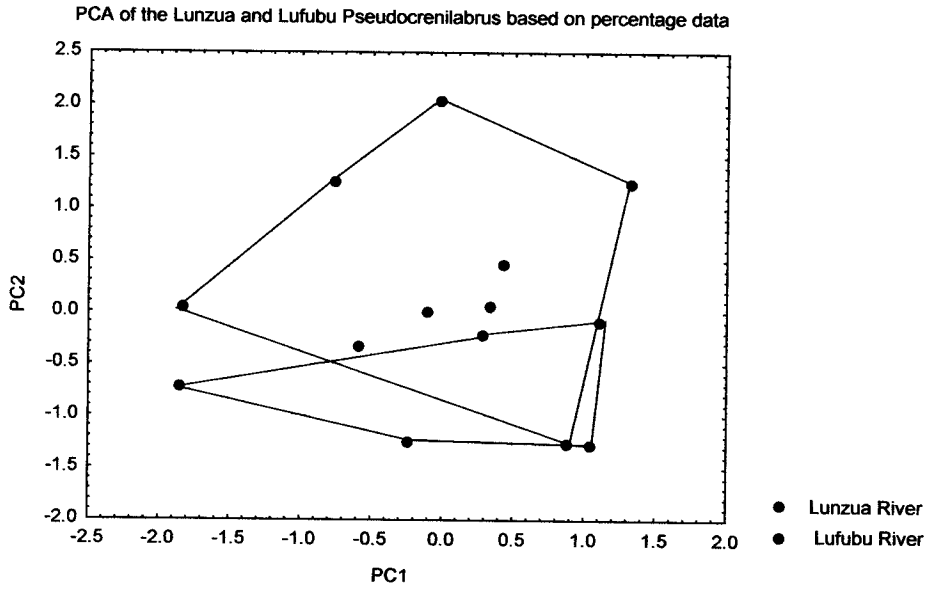
ariable	PC1	PC2	PC3	PC4	PC5
JT	-0.64528	0.559729	0.089463	0.03767	0.294087
JT	-0.63794	0.52408	0.282698	0.052665	-0.10486
JIR	-0.77931	0.076159	0.009051	-0.11612	0.042265
JIR	-0.743	-0.13527	-0.12292	0.232959	-0.08294
ERATO	**-0.59626**	-0.35966	-0.2198	0.102204	0.043631
PI	**-0.51485**	-0.3384	-0.4233	0.145033	0.307937
SPIN	0.294253	-0.34584	0.213517	0.471774	-0.06696
SOFT	-0.03213	0.005308	0.275356	0.558987	-0.35357
SOFT	0.265938	-0.37619	0.124852	0.597351	0.278901
	0.332441	-0.27917	0.337941	0.318808	0.651582
LL	-0.46621	**-0.49624**	-0.31539	0.305034	-0.23962
LL	-0.2609	**-0.50814**	0.625661	-0.33281	-0.15755
OL	-0.44716	**-0.6889**	0.297662	-0.04169	-0.28477
LLA	**-0.59703**	0.376912	0.210825	0.306793	-0.12429
LLA	0.139099	**0.506781**	-0.10231	0.351829	-0.32163
V	-0.28277	**-0.77382**	-0.22368	-0.23202	0.108525
P	**0.534395**	-0.21485	-0.5402	0.045071	-0.44624
K	-0.20485	0.29756	-0.58625	0.138296	0.215586
roportion of total variation	0.230047	0.183691	0.105655	0.088632	0.076352

igure 15. PCA of the meristics of the Zambian *Pseudocrenilabrus* species complex. This is a C1 x PC2 plot, N = 43. The population from the Mwatishi River/Lake Mweru confluence eparates out on PC1. The population from the Lufubu River drainage separates from the opulations from the Lunzua River, Luapula-Chambeshi and the Zambezi-Kafue drainages on C2. The variables with the highest factor loadings are set in bold. Tooth counts are ignored.



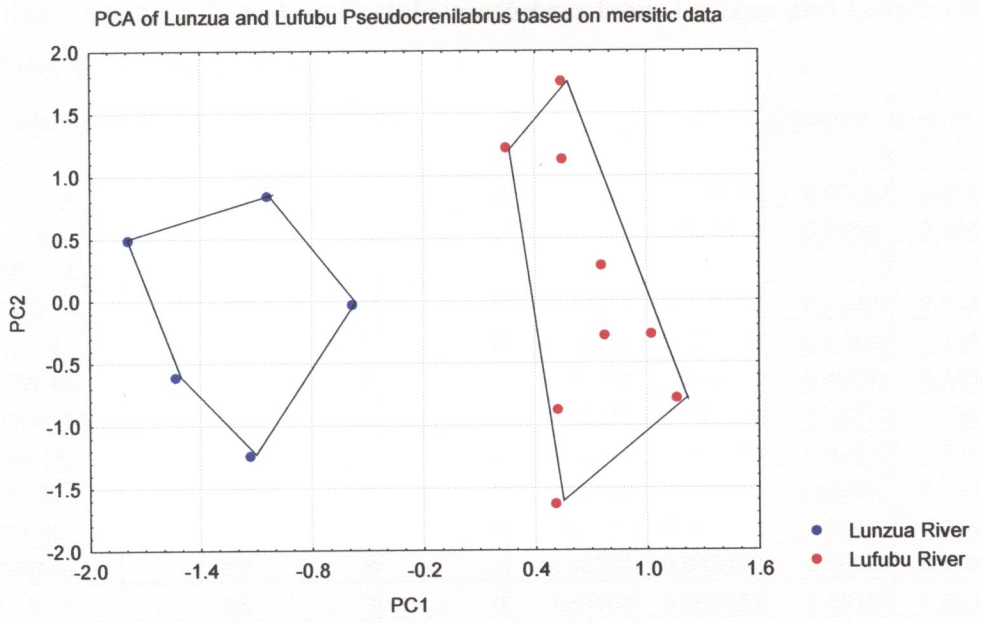
Variable	PC1	PC2	PC3	PC4	PC5
LOGLACRD	0.961554	**-0.19983**	-0.01974	-0.13899	-0.01397
LOGSNL	0.986674	-0.0432	0.06492	-0.09863	0.037828
LOGLJL	0.960701	-0.04827	-0.21559	-0.03182	0.01549
LOGPPL	0.959816	0.17761	-0.12278	-0.06961	0.129752
LOGCHD	0.9603	-0.1156	-0.07388	0.158874	-0.01543
LOGED	0.916036	0.159048	-0.25141	-0.07058	-0.22399
LOGIOW	0.968674	-0.01078	0.088451	-0.10126	0.022532
LOGHW	0.982809	-0.01743	-0.11225	0.002797	-0.00419
LOGHL	0.987188	-0.09328	-0.05394	0.040322	0.092646
LOGSL	0.992539	-0.00381	0.097209	0.052719	-0.00317
LOGBD	0.984756	0.060283	0.092213	-0.09004	-0.02258
LOGDFB	0.959708	**0.175755**	0.137223	-0.09357	-0.00725
LOGAFB	0.896736	**0.225993**	-0.01064	0.370344	0.014836
LOGPRD	0.980189	0.116629	-0.06149	-0.0476	-0.08995
LOGPRP	0.930253	**-0.32599**	0.085459	0.09167	-0.03975
LOGPRV	0.986288	-0.11686	-0.011	0.068185	0.039879
LOGPRA	0.969444	**-0.17486**	0.073695	0.025553	-0.10663
LOGCPL	0.957095	0.075281	-0.0064	-0.04708	0.249009
LOGCPD	0.917157	**0.185779**	0.300796	0.004823	-0.09003
Proportion of total variation	0.924125	0.021755	0.015906	0.013264	0.00904

Figure 16. PCA of *Pseudocrenilabrus* species from the Lunzua and Lufubu Rivers based on log-transformed data. This is a PC2 x PC3 plot with N = 17. There is a separation of the two populations with a major overlap on PC2. The variables with the highest factor loadings on PC2 are set in bold.



Variable	PC1	PC2	PC3	PC4	PC5
LACRD-HL	0.085096	-0.19771	-0.40161	-0.80903	-0.14664
SNL-HL	0.689642	-0.4891	-0.13579	-0.26428	-0.39407
LJL-HL	0.256417	-0.24615	0.550056	-0.53929	0.051928
PPL-HL	0.68539	0.30475	0.50155	0.001773	-0.13526
CHD-HL	0.193539	**-0.85709**	0.295554	-0.01282	-0.07987
ED-HL	0.221874	**0.559636**	0.092711	-0.22651	0.633493
IOW-HL	0.624928	0.064488	-0.48125	-0.12878	-0.26768
IOW-HW	0.128331	0.321305	-0.73846	0.143435	-0.35728
HW-HL	0.561293	-0.3328	0.442306	-0.34257	0.182048
HL-SL	-0.72181	0.28092	0.411387	-0.22998	-0.25672
BD-SL	0.309248	**0.723347**	-0.12246	-0.466	-0.08656
DFB-SL	0.772014	0.329105	0.08975	-0.14169	-0.13462
AFB-SL	0.144513	-0.10242	0.483512	0.657032	0.287673
PRD-SL	0.346972	**0.425397**	0.44009	-0.47157	0.245592
PRP-SL	-0.82694	-0.11436	-0.35736	-0.238	0.189991
PRV-SL	-0.7981	0.188578	0.260014	-0.29554	0.027833
PRA-SL	-0.57276	-0.17172	-0.32367	-0.56372	0.326999
CPL-SL	-0.23993	**0.785117**	0.265609	0.127463	-0.35597
CPD-SL	0.431285	0.464344	-0.43991	0.220442	0.38001
CPD-CPL	0.507314	-0.21044	-0.52132	0.061847	0.562599
Proportion of total variation	0.266541	0.177703	0.162901	0.134791	0.090198

Figure 17. PCA of *Pseudocrenilabrus* species from the Lunzua and Lufubu Rivers based on percentage data. This is a PC1 x PC2 plot with N = 17. There is a separation of the two populations, with an overlap, on PC2. The variables with the highest factor loadings on PC2 are set in bold.



Variable	PC1	PC2	PC3	PC4	PC5
UJT	-0.76116	-0.02183	0.106974	0.013942	-0.12968
LJT	-0.83151	-0.16972	-0.30236	0.159248	0.337591
UJIR	-0.77934	0.093301	-0.06833	-0.29164	-0.2483
LJIR	-0.03196	0.631879	-0.11039	-0.49903	-0.51307
CERATO	**0.899698**	-0.29288	0.002489	-0.13349	0.021753
EPI	**0.752521**	-0.14083	0.109743	-0.33636	0.110943
DSPIN	0.216072	0.220181	-0.62743	-0.2604	0.318626
DSOFT	0.131639	0.37798	-0.28035	0.539522	0.345785
ASOFT	0.131337	0.60789	-0.63721	0.206986	-0.08623
P	0.525497	0.287978	-0.27695	0.450462	-0.25366
ULL	**0.555272**	0.029594	-0.68327	-0.30183	-0.09582
LLL	-0.17973	-0.86554	-0.3351	0.12929	-0.23941
LOL	0.072914	-0.72587	-0.61285	-0.06	-0.24417
DLLA	**-0.90146**	-0.08607	-0.09396	-0.26844	-0.02473
ALLA	-0.50457	-0.07967	-0.29053	-0.53711	0.470595
PV	**0.834021**	-0.45494	0.080097	0.01978	0.056059
CP	0.64056	0.185331	0.24332	-0.53068	0.149369
Proportion of total variation	0.36067	0.159922	0.129433	0.108778	0.066763

Figure 18. PCA of *Pseudocrenilabrus* species from the Lunzua and Lufubu Rivers based on meristic data. This is a PC1 x PC2 plot with N = 14. There is a clear separation of two populations on PC1. The variables with the highest factor loadings on PC1 are set in bold. Tooth counts are ignored.

ann Whitney U-test on *Pseudocrenilabrus* from Lunzua and Lufubu Rivers

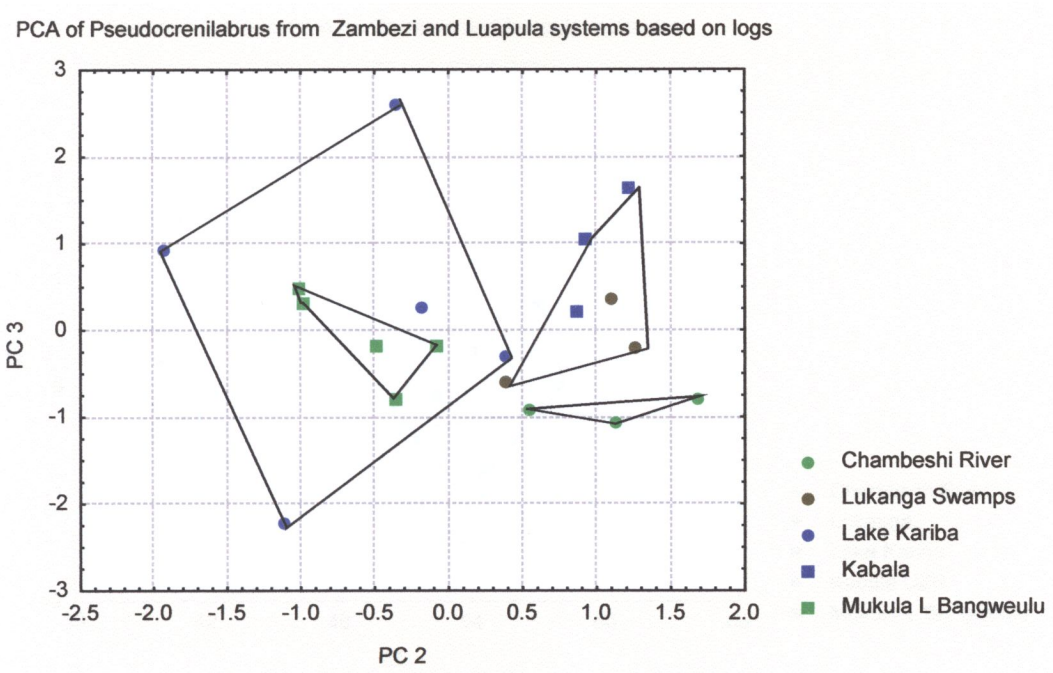
ble 8. Percentage data

riables	Lunzua	Lufubu	U	Z	p-level	Adjusted	p-level	Lunzua	Lufubu
	14	7	4	0	1	0	1	4	2
L-HL	12	9	2	-0.92582	0.354546	-0.92582	0.354546	4	2
-HL	12	9	2	-0.92582	0.354546	-0.92582	0.354546	4	2
L-HL	14	7	4	0	**1**	0	**1**	4	2
ID-HL	15	6	3	0.46291	**0.643432**	0.46291	**0.643432**	4	2
-HL	12	9	2	-0.92582	0.354546	-0.92582	0.354546	4	2
W-HL	13	8	3	-0.46291	**0.643432**	-0.46291	**0.643432**	4	2
W-HW	11	10	1	-1.38873	0.164924	-1.38873	0.164924	4	2
V-HL	16	5	2	0.92582	0.354546	0.92582	0.354546	4	2
-SL	16	5	2	0.92582	0.354546	0.92582	0.354546	4	2
-SL	10	11	0	-1.85164	0.064087	-1.85164	0.064087	4	2
FB-SL	13	8	3	-0.46291	**0.643432**	-0.46291	**0.643432**	4	2
FB-SL	18	3	0	1.85164	0.064087	1.85164	0.064087	4	2
RD-SL	15	6	3	0.46291	**0.643432**	0.46291	**0.643432**	4	2
RP-SL	12	9	2	-0.92582	0.354546	-0.92582	0.354546	4	2
RV-SL	15	6	3	0.46291	**0.643432**	0.46291	**0.643432**	4	2
PD-SL	12	9	2	-0.92582	0.354546	-0.92582	0.354546	4	2
PD-CPL	11	10	1	-1.38873	0.164924	-1.409	0.158844	4	2

able 9. Meristics

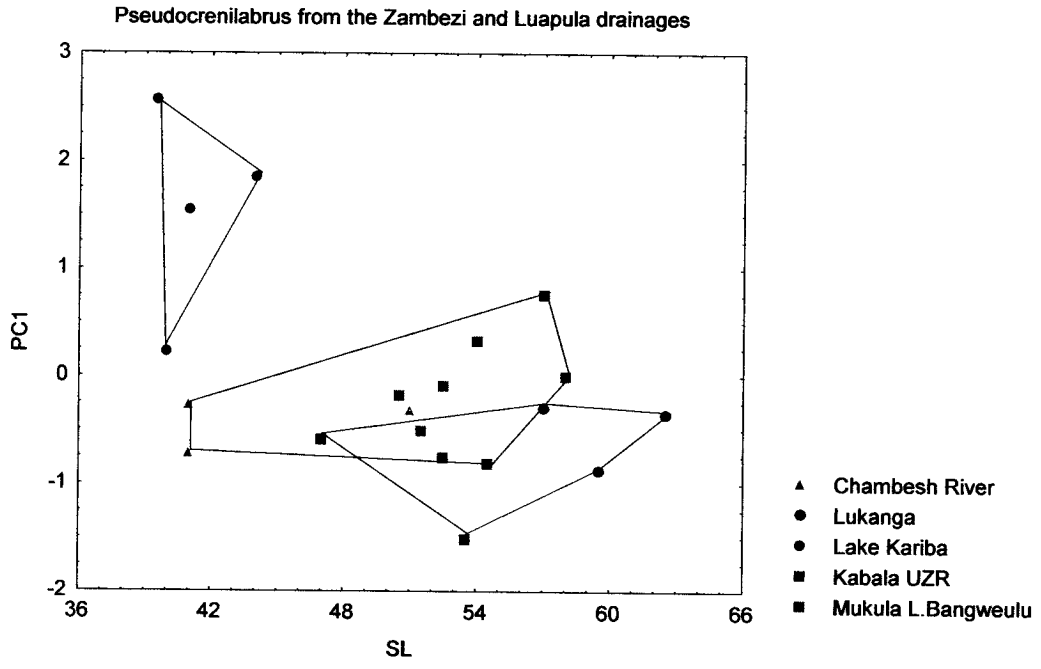
riables	Lunzua	Lufubu	U	Z	p-level	Adjusted	p-level	Lunzua	Lufubu
	14	7	4	0	1	0	1	4	2
JT	16.5	4.5	1.5	1.157275	0.247169	1.17417	0.240335	4	2
IT	15.5	5.5	2.5	0.694365	0.487458	0.704502	0.481125	4	2
IIR	13	8	3	-0.46291	0.643432	-0.55902	0.576154	4	2
ERATO	10	11	0	-1.85164	0.064087	-1.93649	0.052816	4	2
PI	12	9	2	-0.92582	0.354546	-1.41421	0.157309	4	2
SOFT	11	10	1	-1.38873	0.164924	-1.5	0.133624	4	2
SOFT	13	8	3	-0.46291	**0.643432**	-0.70711	**0.479505**	4	2
	12	9	2	-0.92582	**0.354546**	-1.11803	**0.263561**	4	2
LL	10	11	0	-1.85164	0.064087	-1.96748	0.049137	4	2
LL	12.5	8.5	2.5	-0.69437	**0.487458**	-0.75	**0.45326**	4	2
LLA	17	4	1	1.38873	0.164924	1.581139	0.113856	4	2
LLA	15	6	3	0.46291	**0.643432**	0.559017	**0.576154**	4	2
V	11	10	1	-1.38873	0.164924	-1.5	0.133624	4	2
K	14	7	4	0	1			4	2

or both Tables 8 and 9 the p-value for standard length is >0.5 and therefore, the results re valid. The characters with the highest p-values, i.e. contributing the most to the bserved differences between the two populations, are set in bold. Tooth counts are gnored since they are age-related.



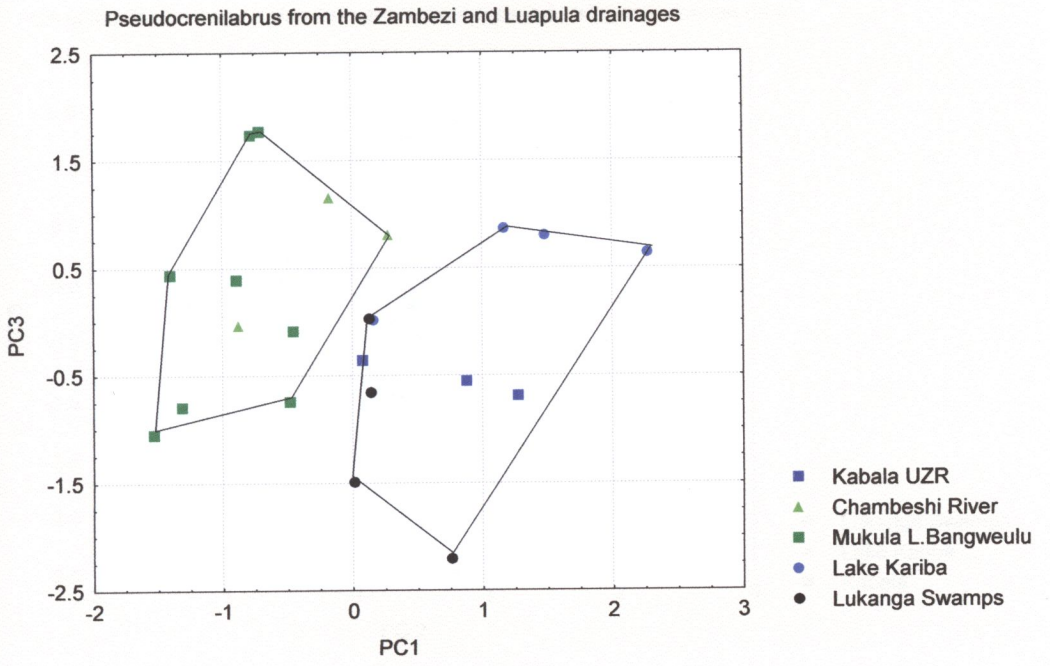
Variable	PC1	PC2	PC3	PC4	PC5
LOGLACRD	0.85417	**-0.22005**	**-0.29044**	-0.29001	-0.02129
LOGSNL	0.855993	**-0.44068**	-0.08903	0.147013	0.0632
LOGLJL	0.937071	0.206337	0.054083	0.10965	0.104084
LOGPPL	0.95191	-0.0856	0.011693	-0.0584	0.083174
LOGCHD	0.936768	0.03788	-0.11416	0.127527	-0.02269
LOGED	0.953925	0.007436	0.078992	0.00067	0.113839
LOGIOW	0.90483	0.10079	-0.17385	-0.18668	0.253847
LOGHW	0.960353	-0.08421	-0.10324	0.133177	-0.08748
LOGHL	0.969786	-0.13883	0.023674	0.1343	0.050138
LOGSL	0.987785	0.025196	0.068451	-0.02823	-0.03485
LOGBD	0.9359	0.1138	**-0.22693**	-0.03792	-0.00284
LOGDFB	0.938497	**0.229127**	-0.08084	-0.05855	-0.13878
LOGAFB	0.852913	**0.404067**	-0.1285	0.200599	-0.01232
LOGPRD	0.907791	-0.0073	0.14885	-0.10508	-0.29692
LOGPRP	0.947405	-0.0553	0.154241	0.218701	-0.00828
LOGPRV	0.966395	-0.04305	0.100216	0.061739	0.066836
LOGPRA	0.954419	-0.17239	0.128421	0.019344	-0.07123
LOGCPL	0.862569	0.088584	**0.349781**	-0.24306	0.186009
Proportion of total variation	0.858936	0.031774	0.022832	0.021568	0.016284

Figure 19. PCA of the *Pseudocrenilabrus* from the Zambezi-Kafue and Luapula-Chambeshi drainage systems. This is a PC2x PC3 plot, N = 22, based on log-transformed data. The Chambeshi population separates from the Kabala and Lukanga populations on PC3 and from the Mukula-Lake Bangweulu population on PC3. The Kariba population spreads out widely. Variables with the highest factor loadings on PC 2 and PC3 are set in bold.



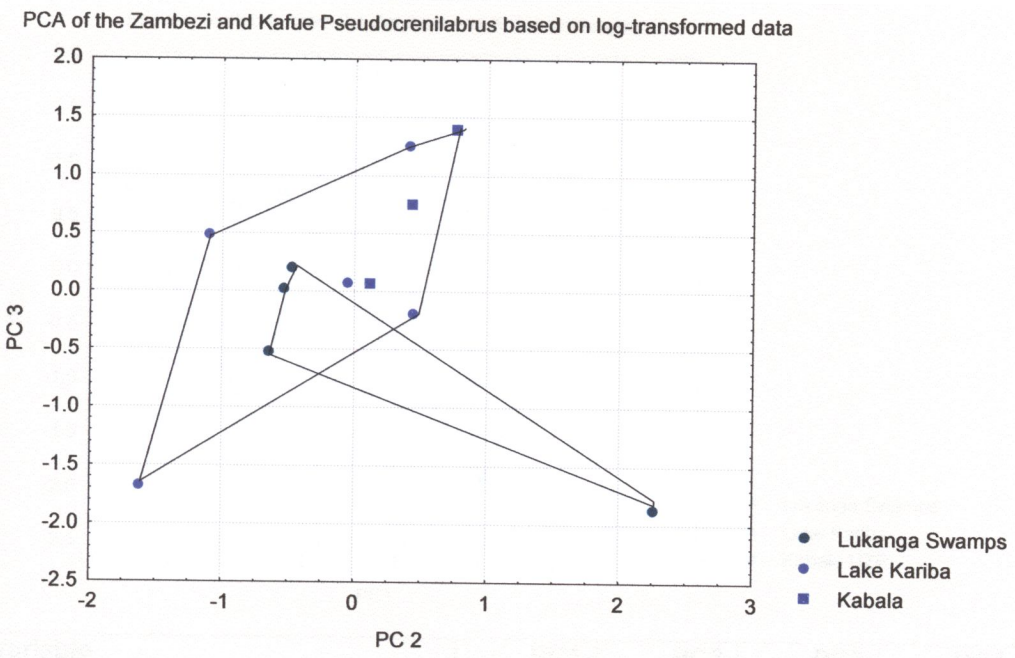
Variable	PC1	PC2	PC3	PC4	PC5
LACRD-HL	-0.15056	-0.16569	-0.38617	-0.73112	0.05803
SNL-HL	**0.537362**	0.107063	-0.2625	-0.39597	0.361825
LJL-HL	**-0.62656**	-0.13572	0.23591	0.41605	0.210119
PPL-HL	-0.53044	0.081609	-0.08506	-0.24602	0.304735
CHD-HL	-0.37734	-0.37961	0.323517	-0.03635	0.344652
ED-HL	-0.24775	-0.07359	-0.19242	0.16228	-0.73507
IOW-HL	**-0.72058**	-0.05368	-0.03298	-0.50072	-0.22672
IOW-HW	**-0.70898**	0.322508	0.175147	-0.4529	-0.19039
HW-HL	0.097746	-0.73232	-0.38237	-0.04677	-0.07151
HL-SL	0.316598	0.361746	0.735631	-0.20394	0.237079
BD-SL	-0.00965	-0.38756	0.529493	-0.39325	0.017359
DFB-SL	-0.33197	-0.73551	0.166613	0.076559	0.071901
AFB-SL	-0.24619	-0.5585	0.246944	0.422098	0.04937
PRD-SL	0.456276	-0.19472	0.004132	0.08607	-0.33146
PRP-SL	0.51626	0.124003	0.518925	0.314793	-0.1394
PRV-SL	-0.38885	0.302189	0.69804	-0.15582	-0.13935
PRA-SL	**0.64599**	0.377762	0.050322	-0.27025	0.029611
CPL-SL	-0.38126	0.726661	-0.09934	0.203665	-0.24582
Proportion of total variation	0.198374	0.162199	0.123067	0.112537	0.083067

Figure 20. PCA of the *Pseudocrenilabrus* from the Zambezi and Luapula drainages. This is a SL x PC1 plot, N = 20, based on percentage data. The Kariba population separates from the others on PC1. For similar-sized individuals, the Lukanga and Kabala populations appear to have lower values on PC1 than the Luapula and Chambeshi populations. The variables with the highest loadings on PC1 are set in bold.



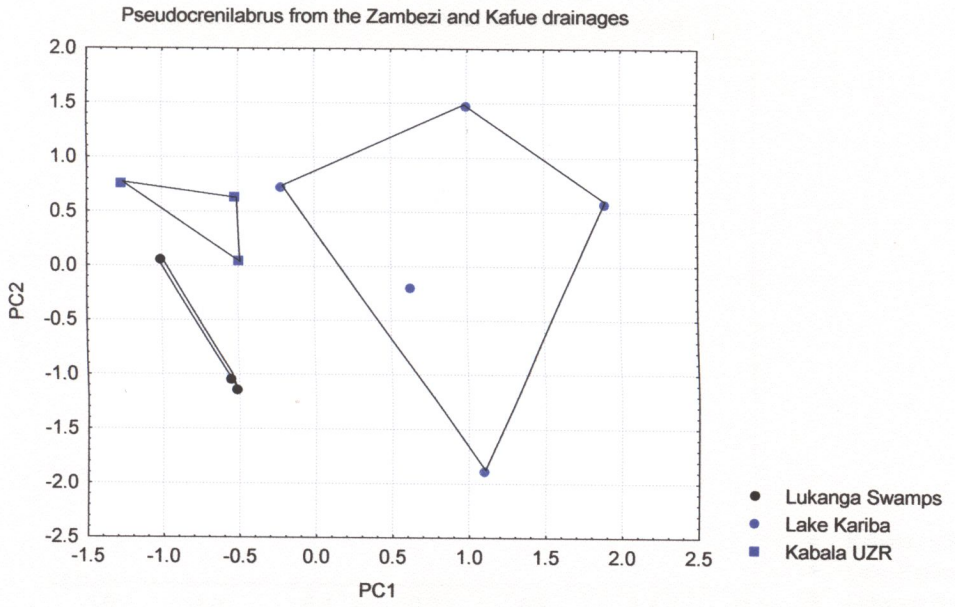
Variable	PC1	PC2	PC3	PC4	PC5
UJT	-0.80898	-0.14253	-0.21485	-0.05385	0.183041
LJT	-0.56091	-0.11049	-0.38363	0.630224	0.067745
UJIR	0.137437	0.01713	-0.63286	-0.46668	-0.47252
LJIR	0.085448	0.396319	-0.6211	0.051517	-0.54271
CERATO	-0.39447	0.345452	**-0.3594**	0.190472	-0.10951
EPI	-0.15307	0.472519	-0.32616	-0.27928	-0.09973
DSPIN	0.261184	-0.1398	**0.571806**	-0.373	-0.47989
DSOFT	-0.0982	0.056073	0.235032	0.391597	-0.56765
ASOFT	-0.24976	0.305643	**0.585127**	-0.18478	-0.24148
P	-0.62604	-0.31157	0.168921	-0.44748	-0.28326
ULL	0.230689	0.689868	0.063492	0.454213	-0.04793
LLL	0.306317	-0.787	-0.08632	0.114535	-0.22298
LOL	**0.497631**	-0.18953	-0.07996	0.585394	-0.23513
DLLA	**-0.55794**	0.327838	**0.310284**	0.069927	-0.33403
ALLA	0.043291	0.590547	0.113813	-0.02126	-0.08525
PV	**0.499835**	0.124865	**-0.50397**	-0.55292	0.122216
CP	**0.783888**	0.296697	0.218744	0.101002	0.071032
CK	-0.07783	0.552308	0.190133	-0.22614	0.413779
Proportion of total variation	0.182483	0.152321	0.135187	0.12201	0.09368

Figure 21. PCA of the *Pseudocrenilabrus* from the Zambezi and Luapula drainages. This is a PC1 x PC3, N = 22, based on meristic data. The Luapula-Chambeshi populations separate from the Zambezi-Kafue populations mainly on PC1. The Lukanga population is limited to the lower regions of PC3. The variables with the highest factor loadings on PC1and PC3 are set in bold. Tooth counts are ignored.



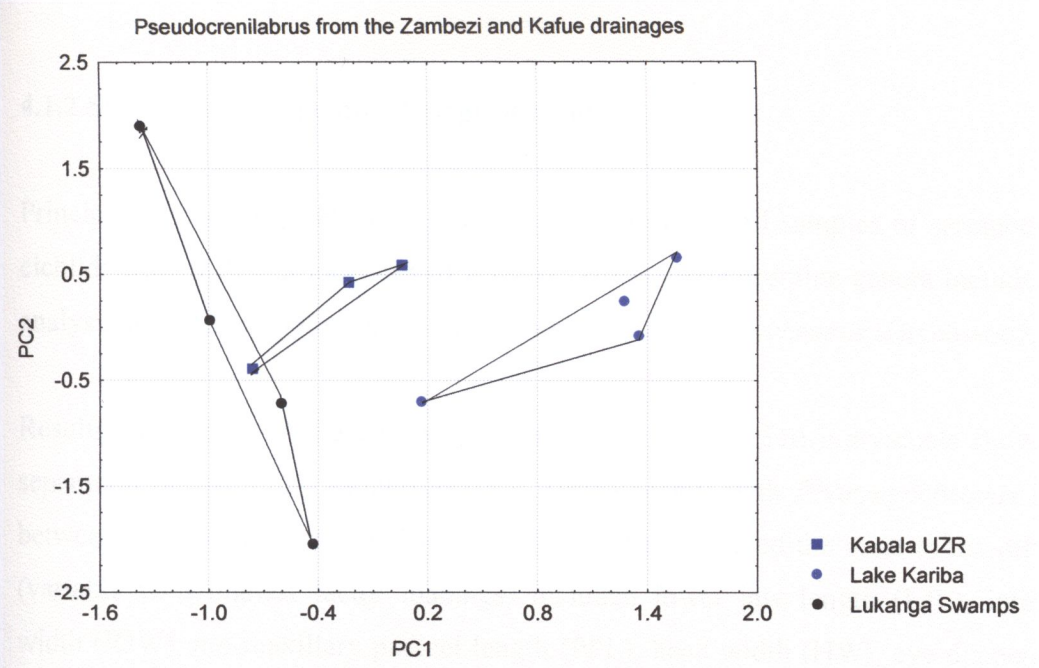
Variable	PC1	PC2	PC3	PC4	PC5
LOGLACRD	0.86924	-0.12101	**-0.31269**	0.006111	0.224387
LOGSNL	0.768454	**-0.59783**	0.00525	-0.18129	0.046044
LOGLJL	0.981706	0.123864	0.039585	0.033285	-0.0582
LOGPPL	0.927791	-0.18745	0.17517	0.21049	0.137401
LOGCHD	0.954928	-0.09275	-0.05363	0.220847	-0.11835
LOGED	0.989825	0.018516	0.041545	-0.02147	0.02402
LOGIOW	0.927459	**0.210378**	-0.14745	0.087466	0.074149
LOGHW	0.892517	0.110021	**-0.34993**	-0.16349	-0.06516
LOGHL	0.967766	-0.19037	0.113104	-0.05257	-0.07565
LOGSL	0.982879	0.131032	-0.0003	-0.07554	0.051878
LOGBD	0.942499	-0.0017	-0.15001	0.251901	-0.12674
LOGDFB	0.964169	0.063894	-0.07878	0.172898	0.026191
LOGAFB	0.980648	0.033092	-0.06418	0.045906	0.064864
LOGPRD	0.907529	-0.1238	0.263523	0.039325	-0.1995
LOGPRP	0.972615	-0.04728	0.106519	-0.1116	-0.00396
LOGPRV	0.947431	-0.10211	0.191368	-0.04241	0.092844
LOGPRA	0.962418	0.089102	0.004881	-0.24184	0.027116
LOGCPL	0.795996	**0.480448**	**0.328565**	-0.04147	0.118368
LOGCPD	0.928368	**0.154806**	-0.10085	-0.17669	-0.20757
Proportion of total variation	0.867906	0.04406	0.029475	0.019636	0.012353

Figure 22. PCA of the *Pseudocrenilabrus* from the Zambezi and Kafue drainages. This is a PC2 x PC3 plot, N = 12, based on log-transformed data. All the populations are similar to each other with the exception of one specimen from the Lukanga population. Variables with the highest loadings on PC2 and PC3 are set in bold.



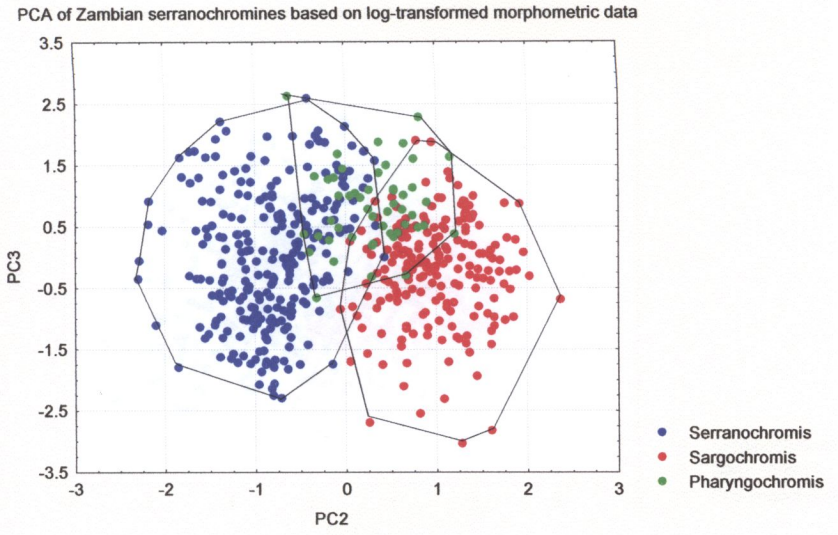
Variable	PC1	PC2	PC3	PC4	PC5
LACRD-HL	-0.21139	-0.12293	0.682361	0.054005	0.299352
SNL-HL	0.631364	-0.01503	0.724562	-0.06547	-0.00688
LJL-HL	**-0.86413**	-0.12917	-0.25909	0.152616	-0.31479
PPL-HL	**-0.89425**	0.06891	0.153666	-0.31087	0.073396
CHD-HL	**-0.72007**	-0.58202	-0.12559	0.053137	0.213985
ED-HL	-0.53293	0.387053	0.182498	0.041672	0.483888
IOW-HL	**-0.83163**	0.022653	0.130468	-0.31967	0.131932
IOW-HW	**-0.84842**	0.052306	-0.2444	-0.22258	0.163108
HW-HL	0.168408	-0.09303	0.855933	-0.22155	-0.03934
HL-SL	0.633862	-0.60801	-0.38849	-0.19413	0.068956
BD-SL	-0.17151	**-0.78296**	0.006557	-0.11809	0.431802
DFB-SL	**-0.75815**	-0.55178	0.092644	0.176949	0.009652
AFB-SL	-0.64039	-0.54704	-0.13136	-0.05591	-0.09383
PRD-SL	0.636991	-0.09988	-0.25889	0.366538	0.519121
PRP-SL	0.690683	-0.19561	-0.3675	-0.41391	0.188363
PRV-SL	0.155255	-0.08209	-0.46533	-0.76635	-0.0941
PRA-SL	**0.907203**	0.127456	-0.06401	-0.168	0.198456
CPL-SL	-0.28551	**0.736061**	-0.50799	0.183398	0.213944
CPD-SL	0.229113	-0.54725	-0.34115	0.487631	-0.07127
CPD-CPL	0.341688	**-0.83878**	0.233603	0.044789	-0.12668
Proportion of total variation	0.381439	0.186984	0.148416	0.080659	0.056992

Figure 23. PCA of the *Pseudocrenilabrus* from the Zambezi and Kafue drainages. This is a PC1 x PC2 plot, N = 11, based on percentage data. The Kariba population separates from the Lukanga and Kabala populations on PC1.The Lukanga and Kabala populations show a close resemblance and separate from each other on PC2. The variables with the highest loadings on PC1 and PC2 are set in bold.



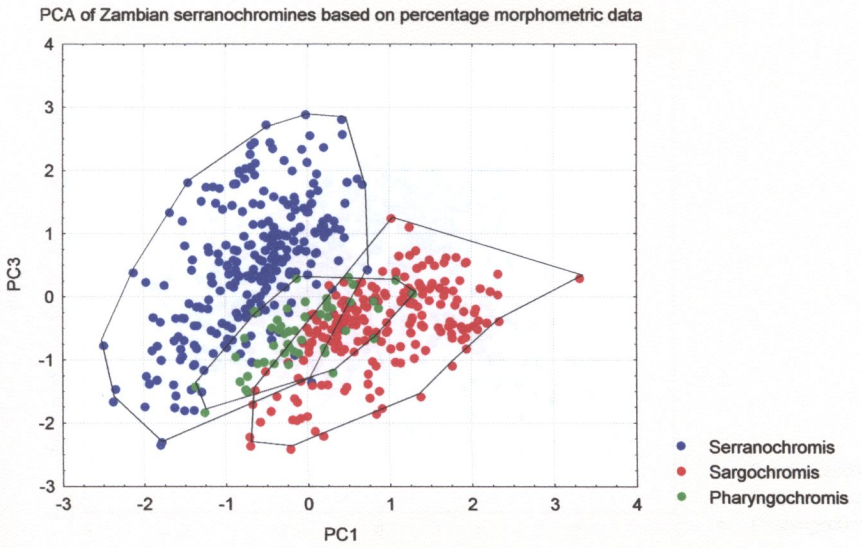
Variable	PC1	PC2	PC3	PC4	PC5
UJT	-0.82717	-0.05776	-0.4263	0.206384	0.129874
LJT	-0.73775	-0.13649	-0.31193	0.154432	-0.11956
UJIR	-0.781	0.181733	0.523625	0.027849	0.235644
LJIR	-0.44382	-0.11479	0.611075	0.448141	-0.092
CERATO	-0.47692	-0.35533	0.559028	-0.03602	-0.31496
EPI	-0.5799	-0.50699	-0.46827	0.268289	0.139604
DSPIN	0.449016	-0.02408	0.463778	0.402358	0.629455
DSOFT	0.382029	-0.67925	-0.07517	0.260611	-0.31844
ASOFT	0.29619	-0.67313	0.264928	0.108623	-0.50726
P	-0.60815	0.080682	0.165144	0.492885	0.462473
ULL	0.304315	-0.6954	-0.12491	-0.03013	0.359443
LLL	0.255584	**0.718755**	-0.13635	0.471182	-0.31398
LOL	0.636688	0.026562	-0.2971	0.501545	0.0517
DLLA	-0.0163	-0.67137	0.545072	-0.10908	-0.04032
ALLA	0.018111	**-0.70172**	-0.23716	0.238301	-0.06207
PV	**-0.74311**	0.319317	0.064471	-0.34816	-0.21278
CP	**0.724997**	0.098156	0.242022	-0.27975	0.366081
CK	-0.29436	-0.60025	-0.28046	-0.47037	0.412503
Proportion of total variation	0.285148	0.210499	0.132989	0.099431	0.098224

Figure 24. PCA of the *Pseudocrenilabrus* from the Zambezi and Kafue drainages. This is a PC1 x PC2 plot, N = 11, based on meristic data. The Lukanga population clearly separates from the Kariba population on PC1. The Kabala population is intermediate. The Lukanga population clearly shows a larger variability on PC1 than the other populations, possibly due to their large size class. Factors with the highest scores on PC1 and PC2 are set in bold. Tooth counts are ignored.



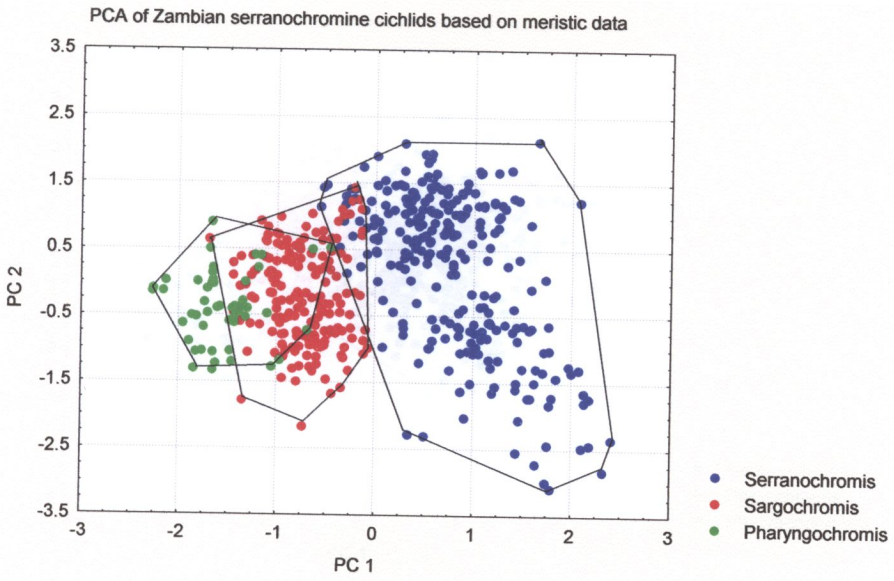
Variable	PC1	PC2	PC3	PC4	PC5
LOGLACRD	0.951124	0.112241	0.174968	0.146406	0.098831
LOGSNL	0.968798	0.030146	0.134664	0.121277	0.079254
LOGLJL	0.929636	**-0.33126**	0.004259	0.009631	-0.05201
LOGPPL	0.943079	**-0.26276**	0.012891	0.092763	-0.04643
LOGCHD	0.981034	-0.05872	0.071293	0.059885	0.00543
LOGED	0.914155	-0.03007	-0.35931	0.024998	0.166742
LOGIOW	0.928826	**0.319477**	0.002234	0.008709	-0.06904
LOGHW	0.94465	**0.212848**	0.030149	-0.02906	0.054654
LOGHL	0.990366	-0.09962	-0.00139	0.007737	-0.01866
LOGSL	0.995631	-0.02677	0.008651	-0.05568	0.002941
LOGBD	0.97227	0.15179	-0.09124	0.017843	-0.07396
LOGDFB	0.969858	0.050839	0.023441	-0.08263	0.029204
LOGAFB	0.965224	0.016859	-0.01411	-0.03666	-0.16945
LOGPRD	0.992416	-0.02157	-0.02272	0.021893	-0.01086
LOGPRP	0.989103	-0.04334	-0.03414	-0.01052	-0.02754
LOGPRV	0.988732	-0.09115	-0.02418	0.003688	-0.01712
LOGPRA	0.994222	0.00531	-0.01856	-0.02945	0.010676
LOGCPL	0.941153	-0.09301	0.152073	-0.23856	0.106022
LOGCPD	0.963487	0.161816	-0.05927	-0.03179	-0.05881
Proportion of total variation	0.930688	0.022425	0.011707	0.006369	0.005677

Figure 25. PCA based on log-transformed data, of the Zambian riverine serranochromines. This plot was between PC2 and PC3. N = 510. There is a clear separation between *Sargochromis* and *Serranochromis* with some overlap on PCA 2. *Pharyngochromis* is intermediate between *Sargochromis* and *Serranochromis* on PC 2. The variables with the highest loadings on PC2 include: LJL, PPL, IOW and HW.



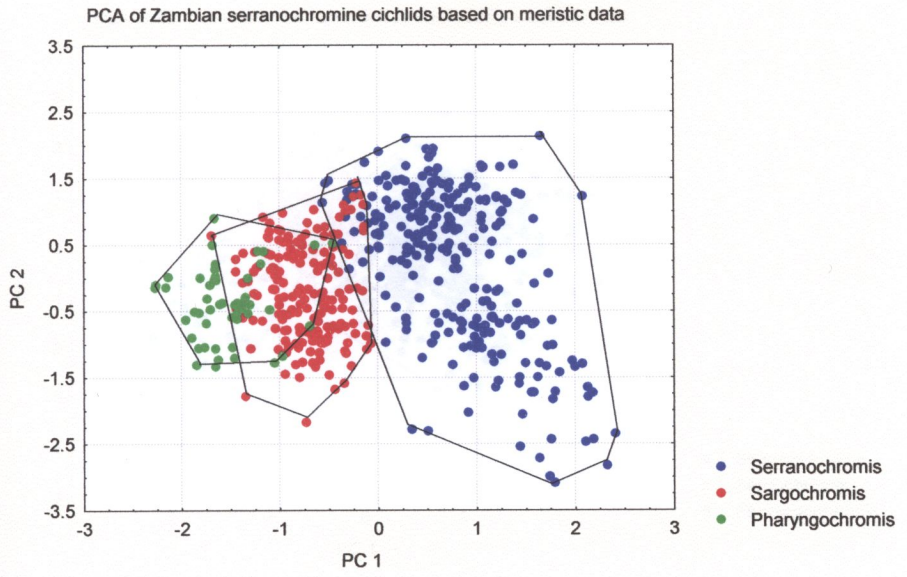
Variable	PC1	PC2	PC3	PC4	PC5
LACRD-HL	0.627679	0.247198	0.328528	-0.352	0.372288
SNL-HL	0.510823	0.237531	0.300333	-0.51763	0.352903
LJL-HL	-0.44759	-0.2281	0.696984	0.037835	-0.25897
PPL-HL	-0.33798	-0.20926	0.566244	-0.05463	0.122453
CHD-HL	0.367351	0.005417	0.754585	-0.16706	-0.00491
ED-HL	-0.12382	0.149202	-0.77927	-0.00254	0.130069
IOW-HL	**0.871624**	0.121554	-0.14202	0.151695	0.179568
IOW-HW	0.377184	-0.13635	0.020198	0.588168	0.539149
HW-HL	**0.689487**	0.300028	-0.19887	-0.34276	-0.26856
HL-SL	-0.4786	-0.68772	-0.00705	-0.15696	0.118101
BD-SL	**0.739598**	-0.41922	-0.10492	0.103659	0.013621
DFB-SL	0.501809	0.082167	0.005339	-0.08274	-0.28397
AFB-SL	0.249026	-0.26441	0.361542	0.365471	0.172463
PRD-SL	0.083682	-0.61761	-0.17513	-0.33215	0.315317
PRP-SL	-0.06904	-0.61012	-0.21598	-0.15549	0.129111
PRV-SL	-0.40919	-0.68032	-0.18115	-0.22478	0.112057
PRA-SL	0.387415	-0.38984	-0.2623	-0.32319	-0.17641
CPL-SL	-0.42797	0.603004	-0.18854	-0.07574	0.30331
CPD-SL	**0.727714**	-0.24071	-0.08521	0.17639	-0.11213
CPD-CPL	**0.739273**	-0.50488	0.018107	0.156079	-0.25854
Proportion of total variation	0.25878	0.156467	0.13026	0.071855	0.060645

Figure 26. PCA based on percentage data, of the Zambian riverine serranochromines. This plot was between PC1 and PC3. N = 510. There is a clear separation between *Sargochromis* and *Serranochromis* with a minor overlap on PC 1. *Pharyngochromis* is intermediate between *Sargochromis* and *Serranochromis* on PC 1. The variables with the highest loadings on PC1include: IOW-HL, HW-HL, BD-SL, CPD-SL and CPD-CPL.



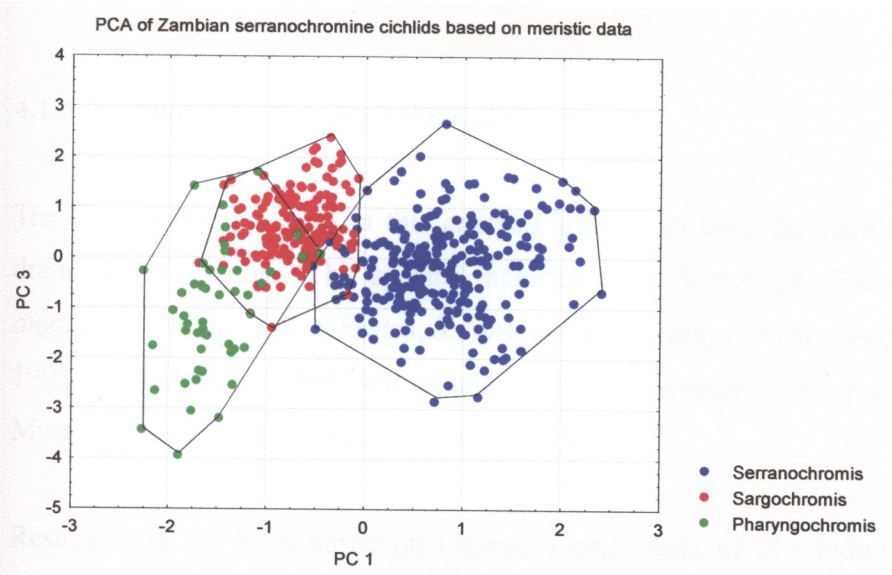
Variable	PC1	PC2	PC3	PC4	PC5
UJT	0.455974	-0.69561	-0.28572	-0.05116	-0.25091
LJT	0.346353	-0.72944	-0.25306	0.12154	-0.28179
UJIR	0.013493	-0.75526	-0.02073	-0.16234	0.401672
LJIR	-0.2664	-0.65349	0.161065	-0.3029	0.38188
CERATO	0.555739	-0.28404	0.351847	0.10655	-0.20031
EPI	0.151991	-0.11426	0.687385	0.318285	0.115523
DSPIN	0.31303	0.017141	-0.23824	0.418722	0.604349
DSOFT	**0.802628**	-0.04786	0.261416	-0.06003	-0.12301
ASOFT	0.774145	-0.08396	0.277964	-0.19879	0.024179
P	0.31727	**0.355349**	0.388067	-0.13312	0.027629
ULL	0.600864	0.076717	-0.22455	0.457958	-0.03165
LLL	0.753309	-0.11591	0.002365	0.228698	0.000748
LOL	**0.835571**	0.013031	-0.14144	0.310294	0.074519
DLLA	0.719501	0.038968	0.024693	-0.28934	0.091753
ALLA	0.55374	0.222461	-0.00799	0.050825	0.372045
PV	0.265945	**0.308465**	-0.465	-0.37496	0.162937
CP	**0.790793**	0.100493	0.032696	-0.26932	-0.01219
CK	**0.799972**	0.099824	0.01245	-0.29039	0.073658
ORPO	**0.83884**	**0.243432**	-0.22247	-0.0874	-0.14819
Proportion of total variation	0.350727	0.13095	0.076807	0.064949	0.057484

Figure 27a. PCA based on meristic data, of the Zambian riverine serranochromines. This plot was between PC1 and PC2. N = 531. Contrary to the results on measurements, there is a clear separation between *Pharyngochromis* and *Serranochromis* on PC 1. In between these two genera, is situated *Sargochromis* with some overlap. The variables with the highest factor loadings on PC1 and PC2 are set in bold.



Variable	PC1	PC2	PC3	PC4	PC5
UJT	0.455974	-0.69561	-0.28572	-0.05116	-0.25091
LJT	0.346353	-0.72944	-0.25306	0.12154	-0.28179
UJIR	0.013493	-0.75526	-0.02073	-0.16234	0.401672
LJIR	-0.2664	-0.65349	0.161065	-0.3029	0.38188
CERATO	0.555739	-0.28404	0.351847	0.10655	-0.20031
EPI	0.151991	-0.11426	0.687385	0.318285	0.115523
DSPIN	0.31303	0.017141	-0.23824	0.418722	0.604349
DSOFT	**0.802628**	-0.04786	0.261416	-0.06003	-0.12301
ASOFT	0.774145	-0.08396	0.277964	-0.19879	0.024179
P	0.31727	**0.355349**	0.388067	-0.13312	0.027629
ULL	0.600864	0.076717	-0.22455	0.457958	-0.03165
LLL	0.753309	-0.11591	0.002365	0.228698	0.000748
LOL	**0.835571**	0.013031	-0.14144	0.310294	0.074519
DLLA	0.719501	0.038968	0.024693	-0.28934	0.091753
ALLA	0.55374	0.222461	-0.00799	0.050825	0.372045
PV	0.265945	**0.308465**	-0.465	-0.37496	0.162937
CP	**0.790793**	0.100493	0.032696	-0.26932	-0.01219
CK	**0.799972**	0.099824	0.01245	-0.29039	0.073658
ORPO	**0.83884**	**0.243432**	-0.22247	-0.0874	-0.14819
Proportion of total variation	0.350727	0.13095	0.076807	0.064949	0.057484

Figure 27a. PCA based on meristic data, of the Zambian riverine serranochromines. This plot was between PC1 and PC2. N = 531. Contrary to the results on measurements, there is a clear separation between *Pharyngochromis* and *Serranochromis* on PC 1. In between these two genera, is situated *Sargochromis* with some overlap. The variables with the highest factor loadings on PC1 and PC2 are set in bold.



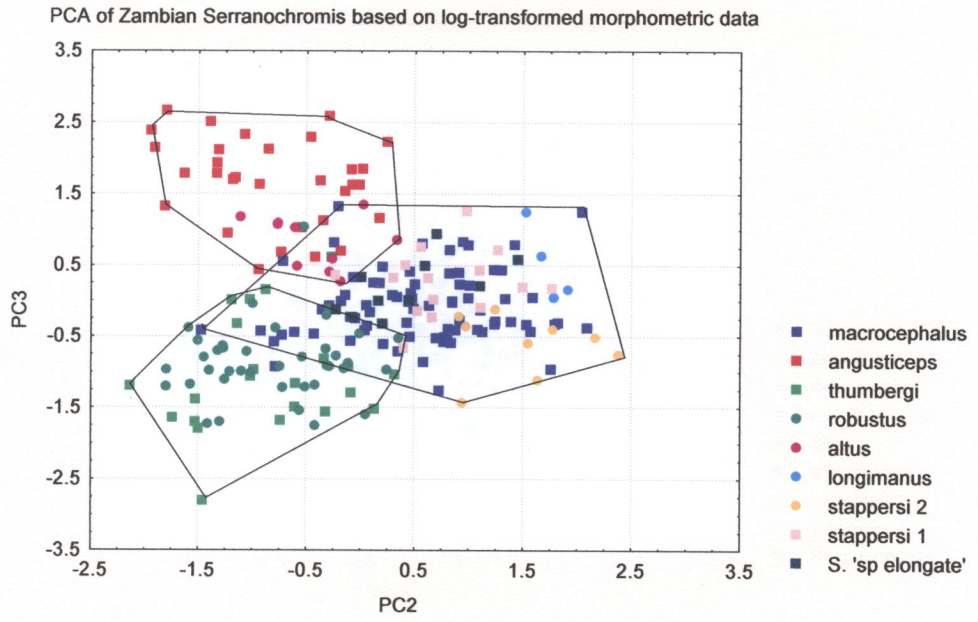
Variable	PC1	PC2	PC3	PC4	PC5
UJT	0.455974	-0.69561	-0.28572	-0.05116	-0.25091
LJT	0.346353	-0.72944	-0.25306	0.12154	-0.28179
UJIR	0.013493	-0.75526	-0.02073	-0.16234	0.401672
LJIR	-0.2664	-0.65349	0.161065	-0.3029	0.38188
CERATO	0.555739	-0.28404	**0.351847**	0.10655	-0.20031
EPI	0.151991	-0.11426	**0.687385**	0.318285	0.115523
DSPIN	0.31303	0.017141	-0.23824	0.418722	0.604349
DSOFT	**0.802628**	-0.04786	0.261416	-0.06003	-0.12301
ASOFT	0.774145	-0.08396	0.277964	-0.19879	0.024179
P	0.31727	0.355349	**0.388067**	-0.13312	0.027629
ULL	0.600864	0.076717	-0.22455	0.457958	-0.03165
LLL	0.753309	-0.11591	0.002365	0.228698	0.000748
LOL	**0.835571**	0.013031	-0.14144	0.310294	0.074519
DLLA	0.719501	0.038968	0.024693	-0.28934	0.091753
ALLA	0.55374	0.222461	-0.00799	0.050825	0.372045
PV	0.265945	0.308465	-0.465	-0.37496	0.162937
CP	**0.790793**	0.100493	0.032696	-0.26932	-0.01219
CK	**0.799972**	0.099824	0.01245	-0.29039	0.073658
ORPO	**0.83884**	0.243432	-0.22247	-0.0874	-0.14819
Proportion of total variation	0.350727	0.13095	0.076807	0.064949	0.057484

Figure 27b. PCA based on meristic data, of the Zambian riverine serranochromines. This plot was between PC1 and PC3. N = 531. The separation on PC1 is similar to that in Fig 21a. There is an additional separation with some overlap between *Sargochromis* and *Pharyngochromis* on PC 3. The variables with the highest factor loadings on PC1 and PC3 are set in bold.

4.1.3 *Serranochromis* morphological results

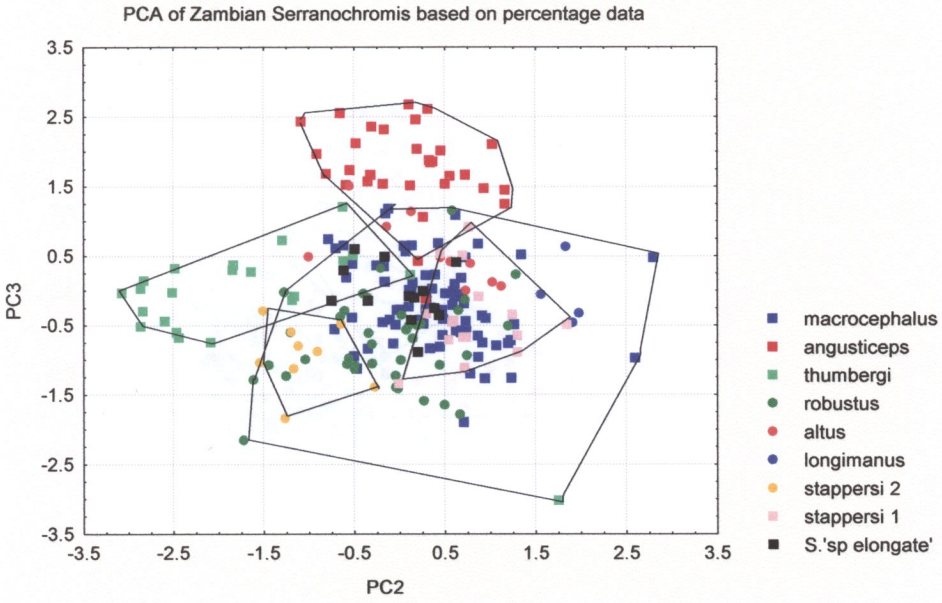
The Serranochromis used in this analysis were from both the Zambezi and the Congo drainages of Zambia. The species included were *Serranochromis macrocephalus*, *S. angusticeps*, *S. altus*, *S. robustus*, *S. longimanus*, *S. stappersi* (Skelton, 2001, Greenwood, 1993 and Trewavas, 1964) and an undescribed *Serranochromis* 'sp elongate' from Lake Mweru, Congo drainage. See Figures 28, 29 and 30.

Results from the PCA based on log-transformed data of 217 individuals cluster, with some overlap, the *Serranochromis* species into three major groups, i.e. the *S. angusticeps* + *S. altus* group, the *S. robustus* + *S. thumbergi* group and the *S. macrocephalus*/ *Serranochromis* 'sp elongate' + *S. stappersi* + *S. longimanus* group. *Serranochromis* 'sp. elongate' is likely to be an undescribed species similar in morphology to *S. stappersi* (Figure 28). The characters that contributed the most to the differences included eye diameter (ED), lachrymal depth (LacrD), snout length (SnL), inter-orbital width (IOW), pre-maxillary pedicel length (PPL) and head width. Figure 29 shows results from the PCA based on percentage data of the 217 individuals. *S. angusticeps* and *S. thumbergi* separate from the rest, with an overlap. *S. stappersi* clearly splits into 2 clusters *with S. stappersi* 1 being similar to *Serranochromis* 'sp elongate'. The characters that contributed the most to the differences included lachrymal depth (LACRD), cheek depth (CHD), lower jaw length (LJL), head length (HL), head width (HW), inter-orbital width (IOW), pre-dorsal distance (PRD), pre-ventral distance (PRV) and caudal peduncle depth (CPD). PCA results based on meristic data group the 245 individuals into three clusters, with overlaps (Figure 30). *Serranochromis* 'sp elongate' appears to be similar to *S. longimanus*. *S. stappersi* 1 and *S. stappersi* 2 clearly separate. The characters that contributed the most to the differences in this analysis were dorsal fin soft rays (DSOFT), anal fin soft rays (ASOFT), ceratobranchial gill rakers (CERATO), lateral line scales (ULL, LOL and LLL), scales between dorsal fin and lateral line (DLLA) and scale rows on the cheek (CK).



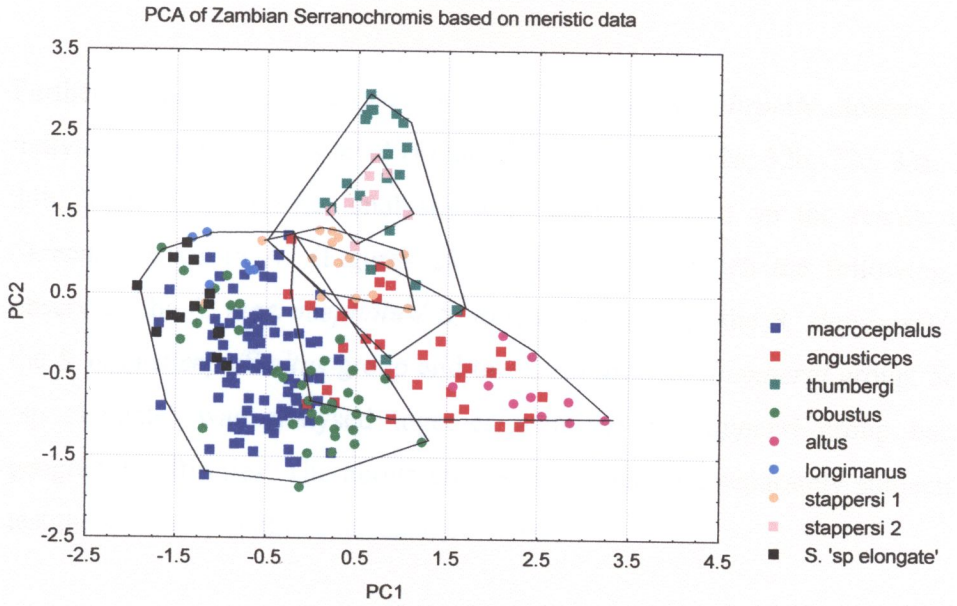
Variable	PC1	PC2	PC3	PC4	PC5
LOGLACRD	0.974569	**-0.15274**	0.055882	0.012914	0.056376
LOGSNL	0.980602	**-0.12158**	0.003417	-0.00118	0.088
LOGLJL	0.983773	-0.00802	0.080808	0.003578	0.054823
LOGPPL	0.961034	0.025496	**0.223547**	-0.04907	-0.05477
LOGCHD	0.985814	-0.09424	0.033452	-0.00352	0.051163
LOGED	0.905914	**0.413659**	-0.02323	0.014886	0.027534
LOGIOW	0.964936	**-0.10668**	**-0.11892**	-0.16617	-0.06112
LOGHW	0.966987	-0.00323	**-0.1668**	-0.09213	0.053612
LOGHL	0.995757	0.011101	0.014566	-0.00383	0.033061
LOGSL	0.996511	0.006523	-0.03134	0.053788	-0.00211
LOGBD	0.986184	0.046072	0.033992	-0.08164	-0.02765
LOGDFB	0.979437	-0.0332	-0.06787	0.068128	-0.03218
LOGAFB	0.971511	-0.04733	0.091016	0.0384	-0.11138
LOGPRD	0.991332	0.030128	0.040502	-0.00399	0.03792
LOGPRP	0.989433	0.03457	0.002261	0.00993	0.028604
LOGPRV	0.993315	0.02256	0.00676	0.012393	0.040665
LOGPRA	0.994432	0.030662	-0.04318	0.024175	0.006677
LOGCPL	0.9596	-0.05709	-0.09142	0.230143	-0.06117
LOGCPD	0.974795	0.029764	-0.0464	-0.06642	-0.13216
Proportion of total variation	0.954212	0.012818	0.006977	0.005939	0.003546

Figure 28. PCA based on log-transformed data, of the Zambian species of genus *Serranochromis*. This plot is between PC2 and PC3. N = 217. *S. altus* and *S. angusticeps* together with *S. robustus* and *S. thumbergi* separate from the rest on PC2 and from each other on PC3. *S. stappersi* splits into two groups. The variables with the highest factor loadings on PC2 and PC3 are set in bold.



Variable	PC1	PC2	PC3	PC4	PC5
LACRD-HL	0.793675	-0.22727	0.262522	0.018608	-0.20443
SNL-HL	0.675399	-0.28912	0.1023	-0.14776	-0.42278
LJL-HL	0.672759	-0.04337	0.303918	-0.22578	0.25287
PPL-HL	0.399253	0.261472	**0.413775**	0.042455	0.381681
CHD-HL	0.826985	-0.1864	0.211626	-0.04298	-0.18032
ED-HL	-0.82161	0.07091	-0.19137	0.100702	0.178053
IOW-HL	0.542234	-0.06747	**-0.5463**	0.282402	-0.35647
IOW-HW	0.167106	0.161782	-0.01208	0.900008	-0.19669
HW-HL	0.369869	-0.2175	**-0.55813**	-0.55757	-0.12374
HL-SL	-0.16899	**0.734721**	0.184769	0.035148	-0.30256
BD-SL	0.470297	**0.630871**	0.002383	0.034965	0.09914
DFB-SL	0.36784	-0.31308	-0.14278	0.237024	-0.00118
AFB-SL	0.498502	0.139613	0.42928	0.0403	0.061135
PRD-SL	-0.00622	**0.682006**	0.296077	-0.1225	-0.13406
PRP-SL	-0.23336	**0.553591**	0.000852	-0.0533	-0.32598
PRV-SL	-0.29442	**0.644506**	0.078712	-0.1689	-0.31721
PRA-SL	0.100756	0.405492	**-0.44421**	-0.36728	-0.0364
CPL-SL	-0.56782	-0.47676	0.103294	0.174121	-0.18301
CPD-SL	0.458312	0.348744	**-0.4805**	0.282927	0.256012
CPD-CPL	0.637863	0.513663	-0.40324	0.104195	0.277424
Proportion of total variation	0.262	0.16714	0.098431	0.082518	0.059137

Figure 29. PCA based on percentage data, of the Zambian species of genus *Serranochromis*. This plot is between PC2 and PC3. N = 217. *S. thumbergi* separates out on PC2 except one individual, while S.*angusticeps* separates out on PC 3. *S. stappersi* splits into two groups. The variables with the highest factor loadings on PC2 and PC3 are set in bold. One *S. thumbergi* is outlying.



Variable	PC1	PC2	PC3	PC4	PC5
UJT	0.717314	0.365507	-0.36884	-0.10856	0.165909
LJT	0.633751	0.56227	-0.32838	0.033911	0.146852
UJIR	0.74728	0.102516	-0.05864	-0.12456	0.069645
LJIR	0.561356	-0.08981	-0.26969	0.3034	0.30472
CERATO	**0.643008**	0.226654	-0.19779	0.278149	-0.12015
EPI	0.410165	-0.28767	0.055218	0.1389	-0.48099
DSPIN	0.211316	0.106054	0.594081	0.142852	0.18964
DSOFT	**0.712923**	-0.14162	-0.12861	0.241544	-0.03278
ASOFT	**0.629543**	**-0.53856**	-0.08547	0.044335	-0.02313
P	-0.16286	-0.29414	0.13961	0.817745	0.153327
ULL	0.245511	**0.546153**	0.494744	-0.11535	-0.00239
LLL	**0.630117**	0.078533	0.324691	-0.05837	-0.11904
LOL	**0.617167**	0.306451	0.586353	-0.02545	-0.02247
DLLA	**0.426774**	**-0.48621**	0.108188	-0.44617	0.166792
ALLA	0.140986	-0.17092	0.582035	0.148819	-0.00375
PV	-0.21797	-0.11324	0.115936	-0.09047	0.771115
CP	0.378604	**-0.472**	0.034336	-0.03719	0.27241
CK	0.422754	**-0.68563**	0.026554	-0.2583	-0.12786
Proportion of total variation	0.264277	0.132222	0.101016	0.071192	0.065631

Figure 30. PCA based on meristic data, of the Zambian species of the genus *Serranochromis*. This plot is between PC1 and PC2. N = 245. *S.macrocephalus, Serranochromis* 'sp. elongate', *S. robustus* and *S. longimanus* with *S. altus* and *S. angusticeps* form two clusters on PC1. *S. thumbergi* and *S. stappersi* separate out together on PC2. *S. stappersi* splits into two groups. The variables with the highest factor loadings on PC2 AND PC3 are set in bold. Tooth counts ignored.

Further analyses were carried out within the *Serranochromis* clusters to separate the individuals up to species level (Figures 31a, 31b, 31c, 32a, 32b, 32c, 33a, 33b, 33c, 34a, 34b, 34c). These further analyses were planned based on the results of the overall *Serranochromis* PCAs (Figures 28, 29 and 30) in which the following clusters were observed: the *S. macrocephalus/ S. longimanus* group, the *S. stappersi1/ S. stappersi2*, the *S. angusticeps/S.altus* group and the *S. robustus/S. thumbergi* group. *Serranochromis* 'sp elongate' was analysed together with the *S. stappersi* group because of their geographic affinities. The above groups were therefore separately subjected to PCAs to test the similarity within each such clusters.

The first analyis involved 52 individuals of *S. macrocephalus* and four (4) individuals of *S. longimanus*. These were subject to PCA using log-transformed data, percentage data and meristic data. In all the analyses there was a separation of the two species, with little overlap and with the exception of one *S. macrocephalus* individual which was remote in all the three analyses. The variables with the highest factor loadings in the log-transformed analysis were eye diameter (ED), pre-maxillary pedicel length, inter-orbital width (IOW), head width (HW), anal fin base (AFB) and pre-dorsal distance (PRD) (Figure 31a). In the percentage data analysis, the variables with the highest loadings were inter-orbital width (IOW), eye diameter (ED), cheek depth (CHD), lachrymal depth (LacrD), body depth (BD), dorsal fin base (DFB), pre-anal distance (PRA), caudal peduncle length (CPL) and caudal peduncle depth (CPD) (Figure 31b). In the meristic data analysis, the variables with the highest factor loadings were scales around the caudal peduncle (CP), scale rows on the cheek (CK), scales between anal fin and lateral line (ALLA), scales between dorsal fin and lateral line (DLLA), lateral line scales (ULL and LLL) and dorsal fin spiny rays (DSPIN) (Figure 31c).

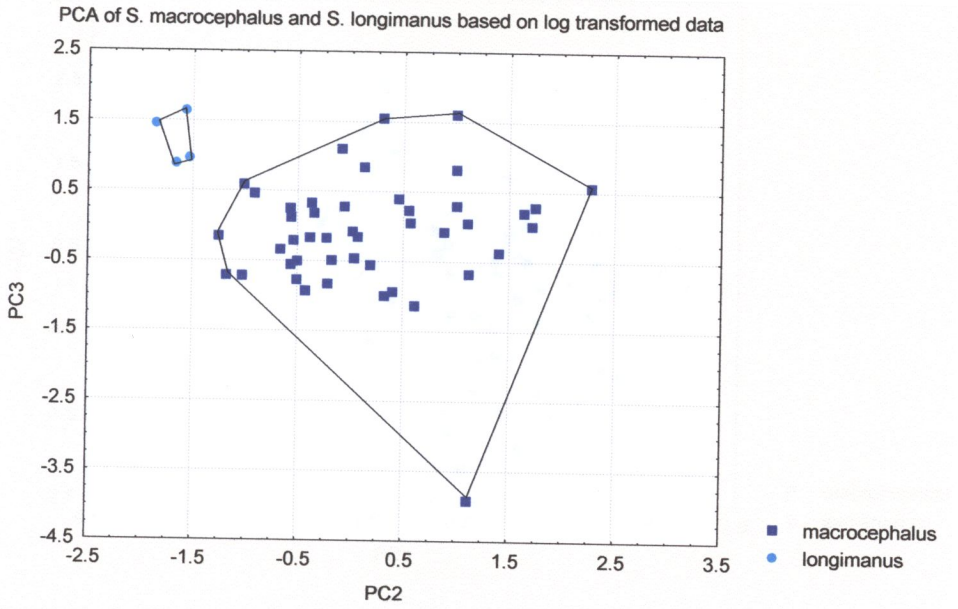
21 individuals of *S. stappersi* and 11 individuals of *Serranochromis 'sp elongate'* were subjected to PCA using the same approach above. All the three analyses showed a clear separation into two groups with small overlaps. The variables with the highest factor loadings in the log-transformed PCA were lachrymal depth (LacrD), snout length (SNL), eye diameter (ED), inter-orbital width (IOW), head width (HW), body depth (BD), anal

fin base (AFB), and caudal peduncle length (CPL) (Figure 32a). In the percentage data PCA, the variables with the highest loadings were the ratios; lachrymal depth (LacrD-SL), snout length (SL-HL), eye diameter (ED-HL), head width (HW-HL), head length (HL-SL), pre-dorsal distance (PRD-SL), pre-pelvic distance (PRP-SL) and pre-ventral distance (PRV-SL) (Figure 32b). The variables with the highest factor loadings in the meristic data analysis were ceretobranchial gill rakers (CERATO), epibranchial gill rakers (EPI), dorsal fin soft and spiny rays (DSOFT and DSPIN), anal fin soft rays (ASOFT), lateral scales (ULL, LOL and LLL), scales between pelvic and ventral fins (PV) and scale rows on the cheek (CK) (Figure 32c).

30 individuals of *S. angusticeps* and six individuals of *S. altus* were subject to PCA using log-transformed data, percentage data and meristic data. In all the three analyses, the two groups separated from each other with some overlaps. The variables with the highest factor loadings in the log-transformed data PCA were lower jaw length (LJL), eye diameter (ED), inter-orbital width (IOW), anal fin base (AFB), pre-ventral distance (PRV) and caudal peduncle length and depth (CPL and CPD) (Figure 33a). In the percentage data PCA, the variable with the highest factor loadings were lachrymal depth (LacrD), cheek depth (CHD), inter-orbital width (IOW), body depth (BD), dorsal fin base (DFB), pre-pelvic distance (PRP), pre-ventral distance (PRV) and caudal peduncle depth (CPD) (Figure 33b). The variables with the highest factor loadings in the meristic data PCA were ceratobranchial gill rakers (CERATO), dorsal fin spiny rays (DSPIN), dorsal fin soft rays (DSOFT), anal fin soft rays (ASOFT), lower lateral line scales (LLL), scales between the lateral line and both the anal and dorsal fins (ALLA and DLLA), scales between pelvic and ventral fins (PV), scales around the caudal peduncle (CP) and scale rows on the cheek (CK) (Figure 33c).

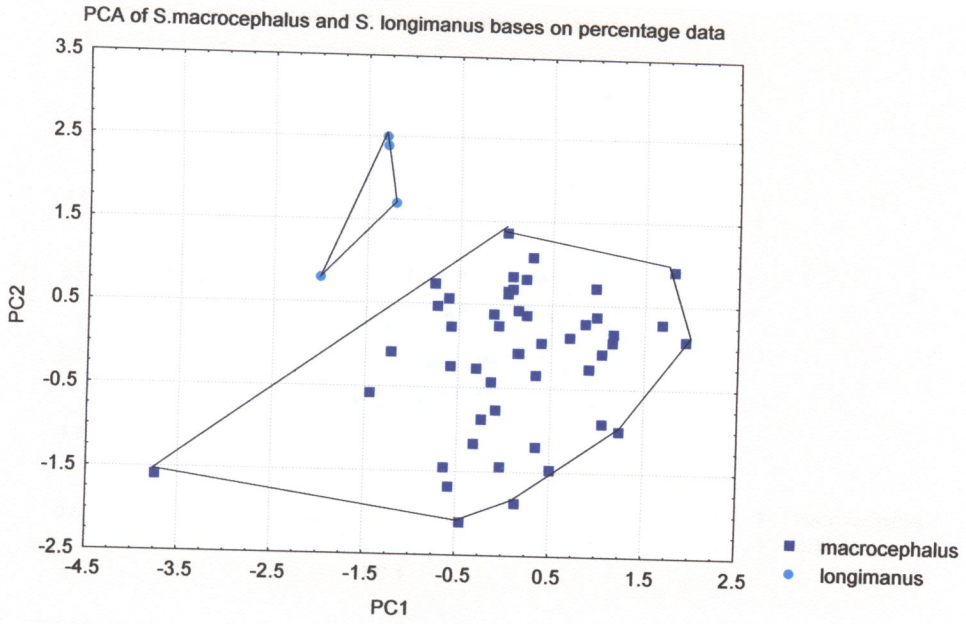
Finally, a total of 39 individuals of *S. robustus* and 20 individuals of *S. thumbergi* were subjected to PCA using log-transformed data, percentage data and meristic data. In all cases there was good separation of the two species. The variables with the highest factor loadings in the log-transformed PCA were pre-maxillary pedicel length (PPL), eye diameter (ED), inter-orbital width (IOW), head width (HW), dorsal fin base (DFB) and

caudal peduncle length (CPL) (Figure 34a). In the percentage data PCA, variables with the highest factor loadings were lachrymal depth (LacrD), cheek depth (CHD), inter-orbital width (IOW), head length (HL), body depth (BD), pre-dorsal distance (PRD), pre-pelvic distance (PRP), pre-ventral distance (PRV), and caudal peduncle length and distance (CPL and CPD) (figure 34b). In the PCA based on meristic data, variables with the highest factor loadings were dorsal fin soft rays (DSOFT), anal fin soft rays (ASOFT), pectoral fin rays (P), lateral line scales (ULL, LOL and LLL), scales between dorsal fin and lateral line (DLLA), scales around the caudal peduncle (CP) and scales rows on the cheek (CK) (Figure 34c).



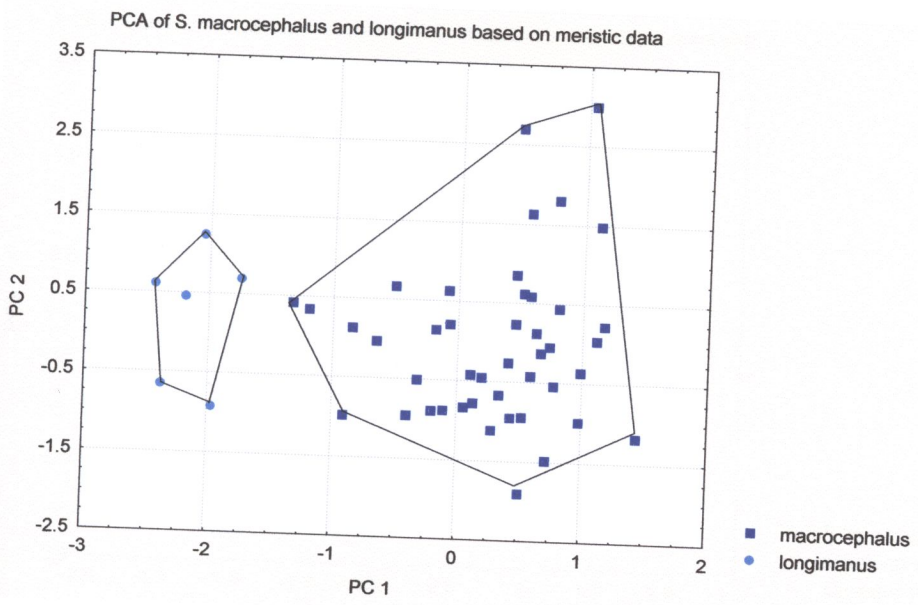
Variable	PC1	PC2	PC3	PC4	PC5
LOGLACRD	0.981676	0.09145	0.064073	0.020888	-0.02398
LOGSNL	0.993211	0.020354	0.036629	0.012201	-0.03574
LOGLJL	0.980499	0.015089	0.043376	0.116378	0.066592
LOGPPL	0.981257	**0.018562**	0.022301	0.087876	0.019037
LOGCHD	0.980044	0.077479	-0.00253	-0.03742	0.035779
LOGED	0.868092	**-0.45968**	-0.00646	-0.11562	0.135726
LOGIOW	0.942552	**0.268324**	**0.081023**	-0.11757	0.06373
LOGHW	0.956319	**0.182051**	0.051146	-0.08052	0.144022
LOGHL	0.994775	-0.03558	0.049619	0.011893	0.004208
LOGSL	0.996376	-0.03038	-0.01822	0.026213	-0.01092
LOGBD	0.987456	-0.02997	0.008644	-0.02909	-0.09375
LOGDFB	0.985293	0.064311	-0.05084	0.04366	0.008025
LOGAFB	0.93742	0.085856	**-0.29874**	-0.13065	-0.02725
LOGPRD	0.988282	**-0.11703**	0.005903	0.037259	-0.03364
LOGPRP	0.986111	-0.10206	0.042344	-0.01064	-0.08087
LOGPRV	0.991463	-0.08328	0.020348	0.013171	-0.05542
LOGPRA	0.992957	-0.04254	0.042453	-0.0047	-0.02926
LOGCPL	0.955831	0.024152	-0.14234	0.209283	0.062144
LOGCPD	0.977232	0.023068	0.036473	-0.07433	-0.12395
Proportion of total variation	0.946582	0.020019	0.007231	0.006768	0.004827

Figure 31a. PCA based on log-transformed data of *S. macrocephalus* and *S. longimanus*. This plot is between PC2 and PC3. N = 52. The two species separate completely on PC2. The variables with the highest factor loadings on PC2 and PC2 are set in bold.



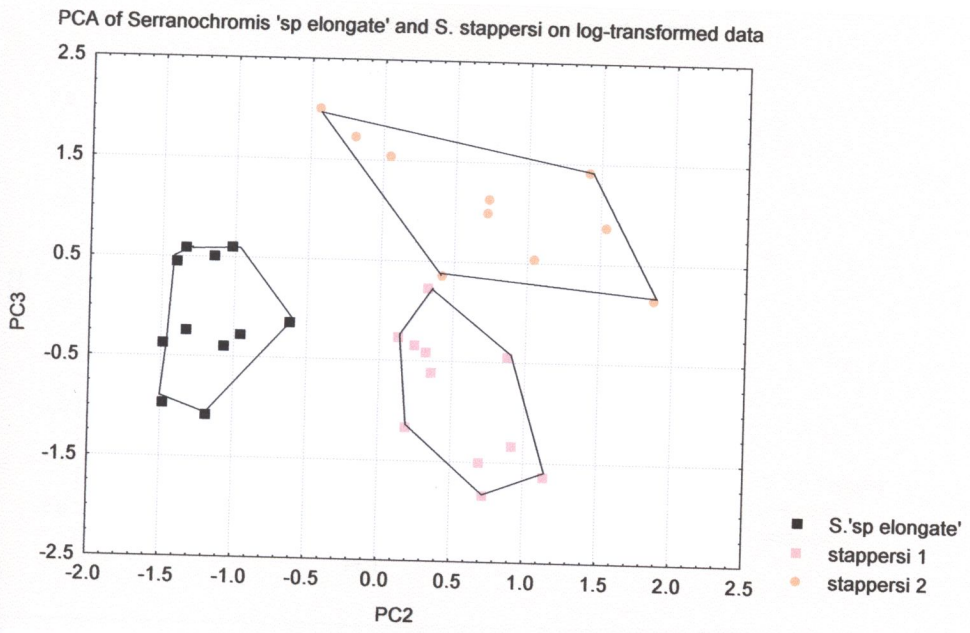
Variable	PC1	PC2	PC3	PC4	PC5
LACRD-HL	**0.727947**	0.314555	-0.104	-0.2196	0.114318
SNL-HL	0.60759	0.287476	-0.37293	0.226722	-0.26229
LJL-HL	0.536269	-0.159	-0.51634	-0.24959	0.27484
PPL-HL	0.312477	-0.18192	-0.63687	-0.05968	0.16061
CHD-HL	**0.730924**	0.056116	0.029724	0.162403	0.124364
ED-HL	**-0.74102**	-0.20781	0.031388	0.336466	-0.23593
IOW-HL	**0.78284**	0.266653	0.094729	-0.12873	-0.1205
IOW-HW	0.500012	0.38873	-0.17756	0.234445	0.36703
HW-HL	0.557695	-0.00922	0.328255	-0.41387	-0.50429
HL-SL	-0.45676	0.372365	-0.00051	-0.59406	0.003332
BD-SL	0.123696	**0.709036**	-0.25191	0.098879	-0.30608
DFB-SL	**0.734847**	-0.17347	0.017722	0.005616	-0.02472
AFB-SL	0.393137	0.023863	-0.01116	0.299594	-0.38262
PRD-SL	-0.63244	0.255365	-0.55668	0.106237	-0.04708
PRP-SL	-0.62238	0.597927	-0.21408	0.031524	-0.01263
PRV-SL	-0.7319	0.390432	-0.20577	-0.10944	-0.12252
PRA-SL	-0.18011	**0.635247**	-0.03638	-0.39736	0.026798
CPL-SL	0.087863	**-0.60697**	-0.47751	-0.14633	-0.46743
CPD-SL	0.315511	**0.749014**	-0.00256	0.130702	-0.27135
CPD-CPL	0.111991	**0.864071**	0.302559	0.18624	0.152415
Proportion of total variation	0.296765	0.19177	0.087753	0.062785	0.061601

Figure 31b. PCA based on percentage data of *S.macrocephalus* and *S. longimanus*. This plot is between PC1and PC2. N = 52. The two species separate completely on PC1 with one *S. macrocephalus* outlying. The variables with the highest factor loadings on PC1 and PC2 are set in bold.



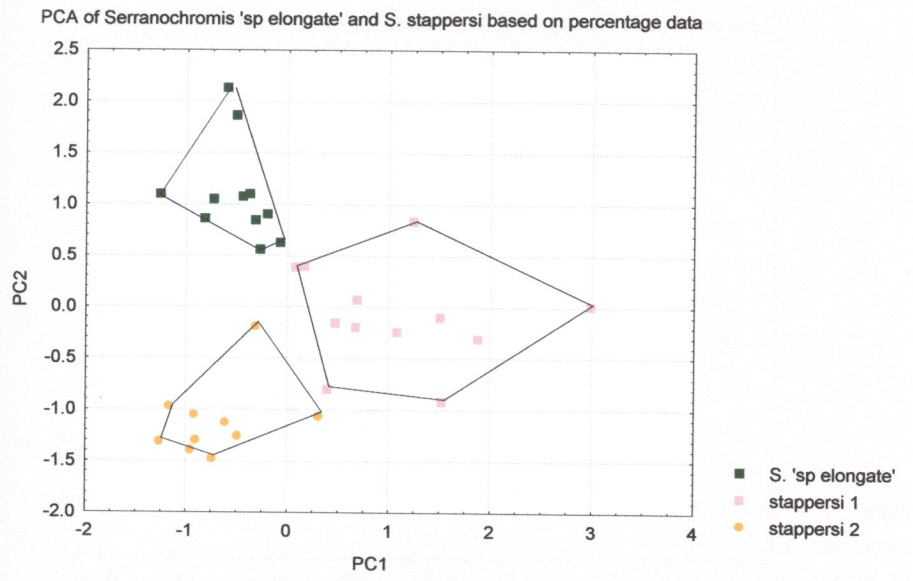
Variable	PC1	PC2	PC3	PC4	PC5
UJT	-0.76212	-0.08985	0.100085	0.172701	-0.14399
LJT	-0.82971	-0.16038	-0.00745	0.109405	0.014655
UJIR	0.217898	-0.23348	0.320781	0.476962	0.231479
LJIR	-0.03168	-0.57681	-0.10254	0.278411	-0.11708
CERATO	-0.32208	-0.27997	-0.10783	0.042397	-0.5199
EPI	0.178169	0.399049	-0.04961	0.576195	0.133846
DSPIN	-0.22078	**-0.61009**	-0.12952	-0.16119	0.027233
DSOFT	0.231706	-0.20445	-0.29925	0.607824	-0.37103
ASOFT	0.38468	-0.09625	-0.45606	-0.10176	-0.29416
P	**0.466562**	0.275347	-0.48048	0.240381	0.244613
ULL	-0.21789	**-0.61857**	0.030117	0.189648	0.548461
LLL	**0.503926**	0.116455	-0.54541	-0.29513	0.025455
LOL	0.344592	-0.48251	-0.51889	-0.03672	0.383029
DLLA	0.377152	**-0.65159**	0.056592	-0.17352	-0.16357
ALLA	**0.630891**	-0.0638	0.472882	-0.09997	0.164664
PV	-0.5246	-0.06839	-0.08416	-0.32249	0.274087
CP	**0.773973**	-0.1145	0.305753	0.046444	-0.02899
CK	**0.695835**	-0.28374	0.389688	-0.12224	-0.2203
Proportion of total variation	0.235564	0.128537	0.096091	0.079433	0.071542

Figure 31c. PCA based on meristic data of *S.macrocephalus* and *S. longimanus*. This plot is between PC1and PC2. N = 50. The two species separate completely on PC1. The variables with the highest factor loadings on PC1 and PC2 are set in bold. Tooth counts are ignored.



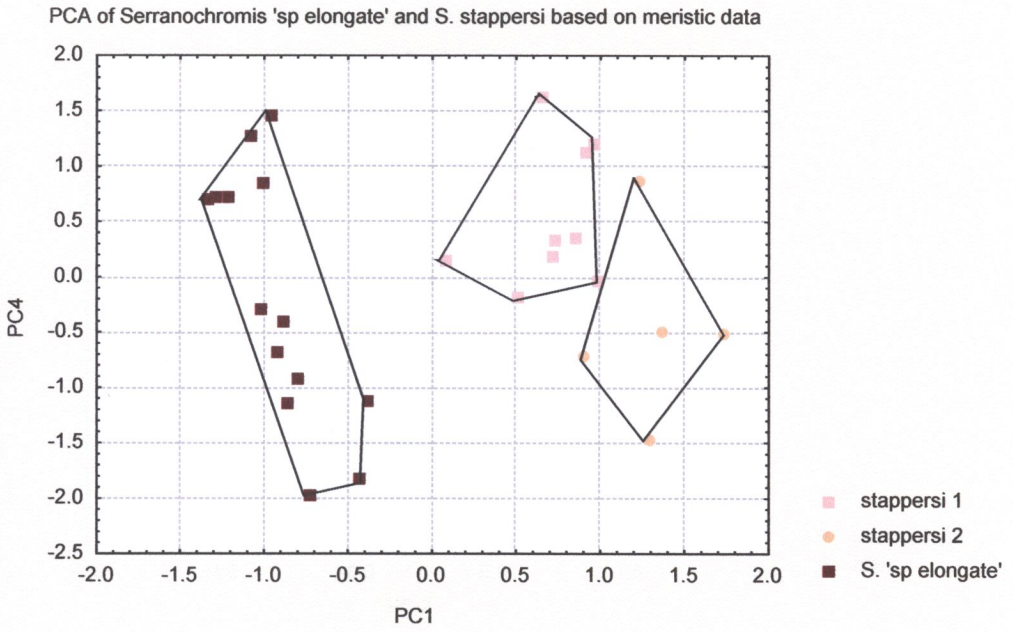
Variable	PC1	PC2	PC3	PC4	PC5
LOGLACRD	0.93527	**-0.30711**	-0.01585	0.01059	-0.01456
LOGSNL	0.951107	**-0.27694**	-0.0692	-0.04556	-0.03059
LOGLJL	0.981063	-0.11119	-0.08325	-0.05516	-0.03493
LOGPPL	0.952189	-0.10801	0.103982	-0.01231	0.066474
LOGCHD	0.963135	-0.16828	-0.04447	-0.02583	-0.01153
LOGED	0.874133	**0.342288**	**0.265686**	-0.11096	0.035083
LOGIOW	0.878497	**0.205239**	**-0.27616**	0.281804	0.109479
LOGHW	0.921891	-0.09774	**0.183117**	0.13002	0.262141
LOGHL	0.98846	-0.06083	-0.0275	-0.03161	-0.03421
LOGSL	0.990822	0.063457	0.089115	0.000579	-0.05455
LOGBD	0.925043	0.136682	**-0.28977**	0.02317	-0.00579
LOGDFB	0.972368	0.158096	0.074282	-0.0295	-0.05372
LOGAFB	0.90492	**0.185144**	-0.15117	-0.26153	0.141601
LOGPRD	0.983534	-0.02201	-0.04629	-0.04675	-0.00331
LOGPRP	0.986436	0.006134	-0.04561	-0.02028	-0.04893
LOGPRV	0.988191	-0.00459	-0.0698	-0.05749	-0.0407
LOGPRA	0.98902	0.008702	0.059482	0.029146	-0.04505
LOGCPL	0.937724	-0.05063	**0.251384**	0.088993	0.021803
LOGCPD	0.937339	0.153399	0.089741	0.148104	-0.22464
Proportion of total variation	0.904934	0.026557	0.021582	0.011689	0.009105

Figure 32a. PCA based on log-transformed data of *S. stappersi* and *Serranochromis* 'sp elongate'. This plot is between PC2 and PC3. N =32. *S. stappersi* 1 and *S. stappersi 2* separate completely from *Serranochromis* 'sp elongate' on PC 2 and from each other on PC3. The variables with the highest factor loadings on PC2 and PC3 are set in bold.



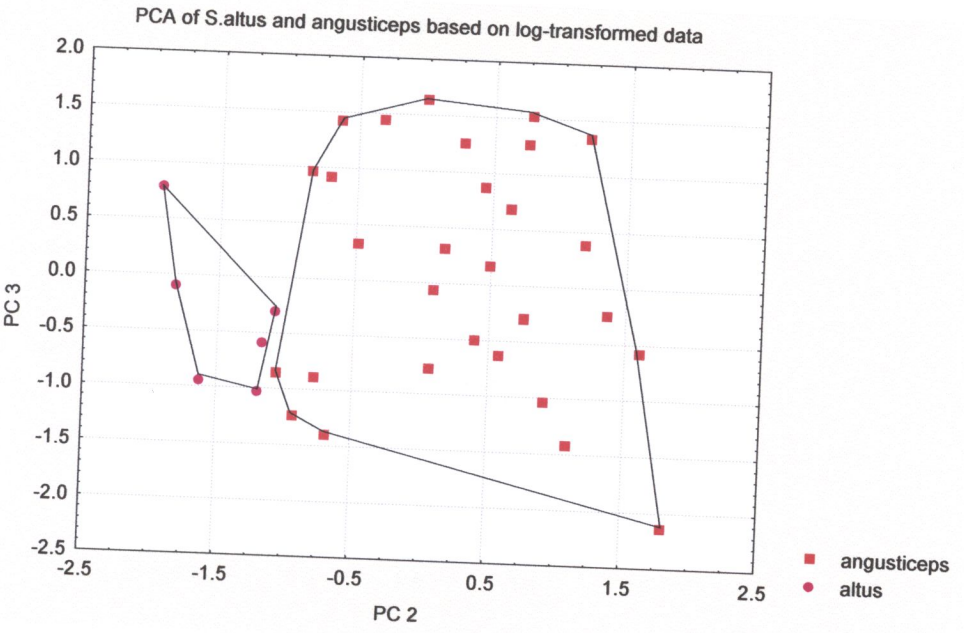
Variable	PC1	PC2	PC3	PC4	PC5
LACRD-HL	**-0.7337**	0.505714	0.187437	-0.11042	0.062247
SNL-HL	-0.63152	**0.618533**	0.303134	-0.0409	0.180134
LJL-HL	-0.55252	0.270884	0.628559	0.164277	0.010053
PPL-HL	-0.61919	-0.07735	0.196924	0.327617	0.148301
CHD-HL	-0.65496	0.373231	0.298557	-0.03907	-0.21627
ED-HL	0.288984	**-0.65482**	-0.34896	0.164786	-0.33152
IOW-HL	0.104722	-0.21337	0.513993	-0.21845	-0.63098
IOW-HW	0.615953	-0.1284	0.454807	-0.39246	-0.38397
HW-HL	**-0.70254**	-0.02902	-0.07402	0.265	-0.06484
HL-SL	**0.673251**	**0.637045**	-0.22076	-0.01358	0.149488
BD-SL	0.565113	0.242181	0.586532	0.160325	-0.00896
DFB-SL	-0.03687	-0.49808	0.450874	0.52617	0.100852
AFB-SL	0.236925	0.18495	0.329392	0.568061	-0.39614
PRD-SL	0.501391	**0.561322**	0.110549	0.161343	-0.11159
PRP-SL	**0.750362**	0.50613	-0.03257	0.036132	0.178648
PRV-SL	**0.786586**	**0.556**	-0.03575	0.067967	0.101949
PRA-SL	-0.35553	0.51795	0.132842	-0.47072	-0.30559
CPL-SL	-0.47787	0.111484	-0.6138	-0.19324	-0.19392
CPD-SL	-0.02754	-0.50605	0.384492	-0.50777	0.358001
CPD-CPL	0.317105	-0.39373	0.712665	-0.21269	0.379182
Proprtion of total variation	0.287078	0.183381	0.150877	0.083411	0.070604

Figure 32b. PCA based on percentage data of *S. stappersi* and *Serranochromis* 'sp elongate'. This plot is between PC1 and PC2. N =32. *S. stappersi* 1 and *S. stappersi* 2 separate completely from *Serranochromis* 'sp elongate' on PC2 and from each other on PC1. The variables with the highest factor loadings on PC1 and PC2 are set in bold.



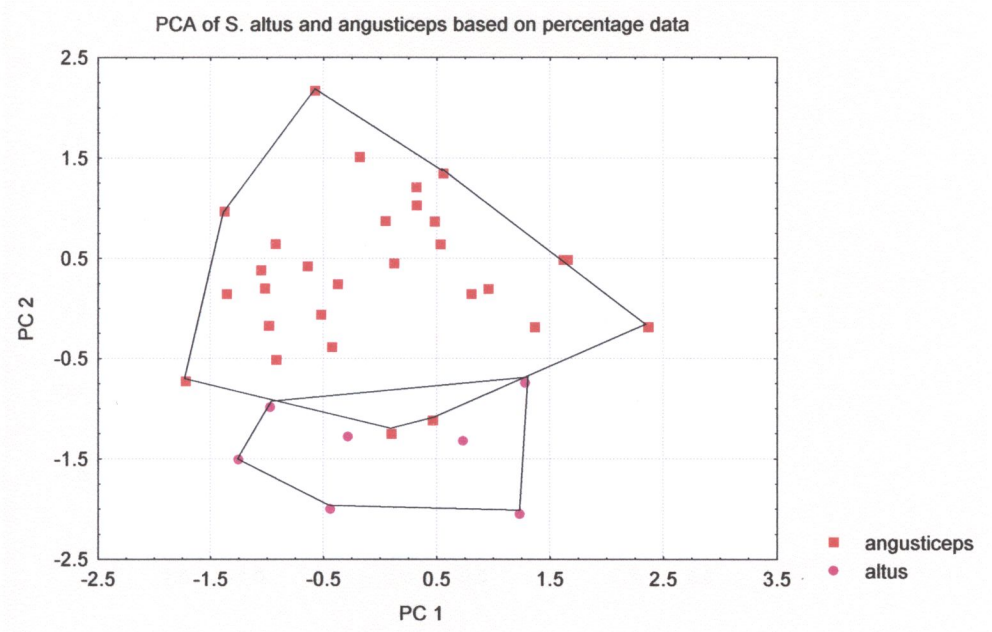
Variable	PC1	PC2	PC3	PC4	PC5
UJT	0.87278	-0.07172	0.182205	0.046103	-0.04463
LJT	0.883804	-0.08883	0.305323	0.140643	-0.08668
UJIR	0.610463	0.323849	-0.47064	0.015318	0.292102
LJIR	0.497941	-0.03408	0.235023	0.183109	0.16793
CERATO	**0.86729**	0.072825	0.198842	0.265725	0.167445
EPI	**0.662524**	0.275584	-0.24548	0.15028	-0.37515
DSPIN	-0.48713	-0.02483	0.436327	0.053453	-0.56957
DSOFT	0.629947	-0.06059	0.20592	**-0.46329**	0.331602
ASOFT	0.210295	0.550806	0.318651	**0.458737**	0.343147
P	-0.30474	-0.14963	0.585736	-0.27663	0.497156
ULL	0.301944	0.130488	-0.11279	**-0.83007**	-0.14242
LLL	**0.663931**	0.423998	-0.23002	**0.056904**	-0.26754
LOL	0.410826	0.526052	-0.25222	**-0.36398**	0.129156
DLLA	0.009848	0.665255	0.465968	-0.08787	-0.34043
ALLA	-0.37864	0.614378	0.486409	-0.31715	0.009903
PV	**-0.65016**	0.534102	0.020135	0.145771	0.034226
CP	0.537439	-0.21183	0.579012	0.071762	-0.10184
CK	**-0.64924**	0.41676	-0.15063	0.244293	0.305589
Proportion of total variation	0.33906	0.129321	0.117871	0.092913	0.078919

Figure 32c. PCA based on meristic data of S. *stappersi* and *Serranochromis* 'sp elongate'. This plot is between PC1 and PC4. N =28. *Serranochromis* 'sp elongate'separates completely from *S. stappersi* 1 and 2 on PC1 .*S. stappersi* 1 and 2 separate from each other with an overlap, on PC 1 and PC4. The variables with the highest factor loadings on PC1 and PC4 are set in bold. Tooth counts ignored.



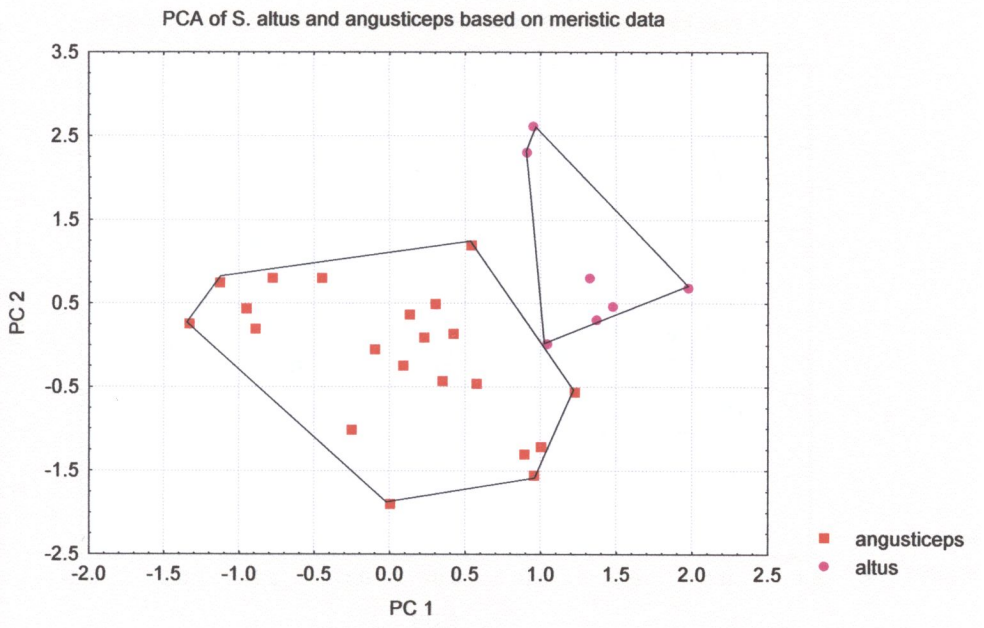
Variable	PC1	PC2	PC3	PC4	PC5
OGLACRD	0.967239	0.140965	-0.11687	0.045078	0.137196
OGSNL	0.985962	0.083737	-0.01965	-0.01921	0.08455
OGLJL	0.981274	0.024545	**0.141361**	0.029012	-0.01453
OGPPL	0.972174	0.080179	0.074422	0.016035	0.12865
OGCHD	0.984525	0.09551	0.032421	0.079785	-0.00188
OGED	0.909328	**-0.34049**	**0.166343**	-0.08699	-0.01832
OGIOW	0.93463	**-0.22591**	-0.22007	-0.04314	0.055273
OGHW	0.976275	-0.12451	-0.027	-0.08862	0.056396
OGHL	0.995722	0.019508	0.058562	0.017077	-0.01137
OGSL	0.994605	0.047257	0.042257	0.004096	-0.04884
OGBD	0.982863	-0.08279	-0.07484	0.103622	-0.02294
OGDFB	0.987817	0.026975	-0.02866	0.025906	-0.04626
OGAFB	0.934716	**0.208215**	**-0.17342**	-0.06853	-0.14002
GPRD	0.98902	0.051277	0.103053	0.038962	0.027249
GPRP	0.991503	-0.01986	0.073888	0.036593	-0.04466
GPRV	0.98491	0.018274	**0.137057**	0.031482	-0.03706
GPRA	0.989699	0.005989	0.05241	0.007695	-0.05147
GCPL	0.953019	**0.116183**	-0.0247	-0.2502	-0.00537
GCPD	0.94819	**-0.15255**	**-0.21584**	0.104646	-0.0487
portion of total variation	0.944884	0.016801	0.01284	0.00637	0.004382

ure 33a. PCA based on log-transformed data of *S. altus* and *S. angusticeps*. This plot etween PC2 and PC3. N = 36. The two species completely separate on PC2. The ables with the highest factor loadings on PC2 and PC3 are set in bold.



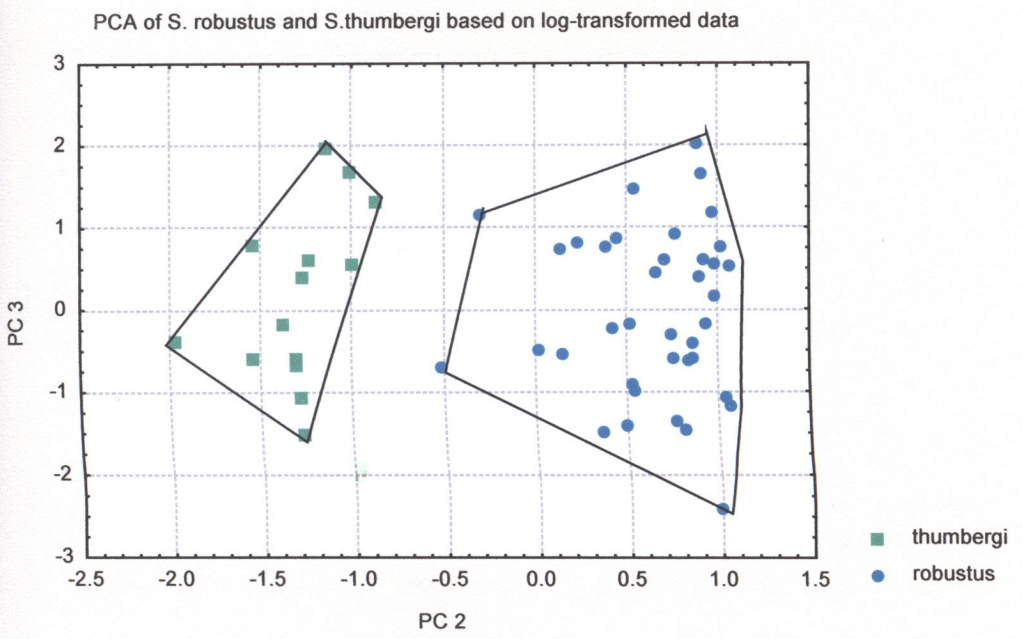
Variable	PC1	PC2	PC3	PC4	PC5
LACRD-HL	**0.643306**	0.091433	0.613216	-0.16925	0.124149
SNL-HL	0.461298	0.183089	0.687759	0.0704	-0.02791
LJL-HL	-0.24617	0.244343	0.010585	-0.38721	-0.55253
PPL-HL	-0.16344	0.190051	0.694468	-0.18889	-0.28639
CHD-HL	**0.632089**	0.237597	-0.01105	-0.49595	-0.07461
ED-HL	-0.62101	-0.42238	0.131918	0.289433	-0.27015
IOW-HL	0.466884	**-0.71234**	0.174897	0.341764	-0.06417
IOW-HW	0.251835	**-0.61394**	0.165863	0.100787	0.362462
HW-HL	0.41041	-0.36912	0.061158	0.419493	-0.54187
HL-SL	-0.66421	-0.38401	0.310988	0.048688	0.140539
BD-SL	0.509161	**-0.67282**	-0.02645	-0.28461	-0.11588
DFB-SL	**0.716115**	-0.20436	-0.16303	-0.12412	0.000822
AFB-SL	0.447689	0.131643	0.394116	0.031853	0.407732
PRD-SL	-0.59093	-0.04029	0.583789	-0.33638	-0.05901
PRP-SL	**-0.68155**	-0.54303	0.094107	-0.05363	-0.10365
PRV-SL	**-0.78831**	-0.17302	0.123347	-0.16638	0.158577
PRA-SL	-0.51534	-0.43754	0.085971	0.25451	0.215339
CPL-SL	0.286132	0.399775	0.198996	0.700897	-0.15304
CPD-SL	0.486347	**-0.76835**	0.039955	-0.03187	-0.13096
CPD-CPL	0.181901	**-0.81055**	-0.11342	-0.4518	-0.01191
Proportion of total variation	0.270743	0.198873	0.105823	0.092828	0.061853

Figure 33b. PCA based on percentage data of *S. altus* and *S. angusticeps*. This plot is between PC2 and PC3. N = 36. The two species separate with a minor overlap on PC2. The variables with the highest factor loadings on PC1 and PC2 are set in bold.



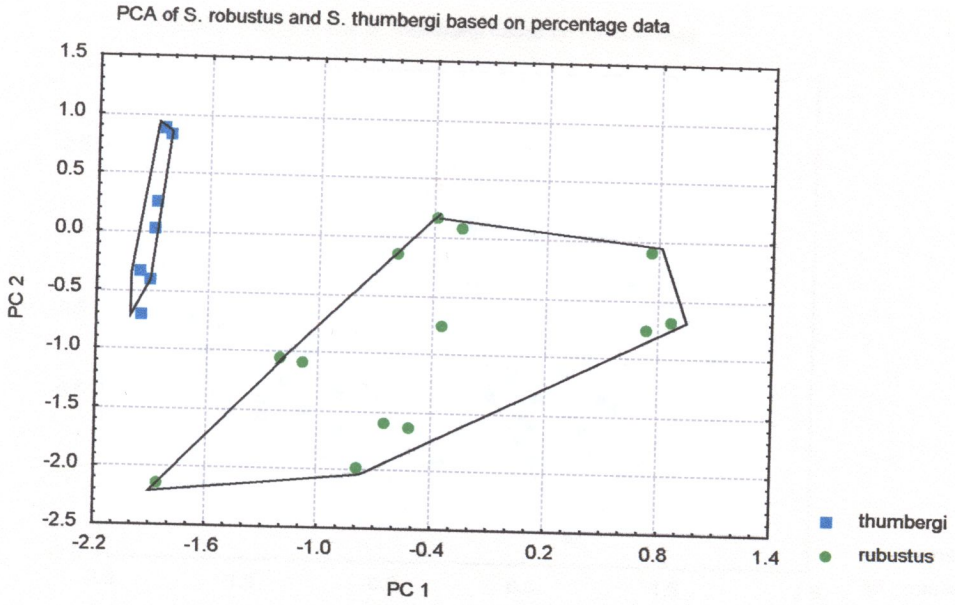
Variable	PC1	PC2	PC3	PC4	PC5
UJT	0.636406	0.528102	-0.30813	0.022992	-0.04945
LJT	0.229959	0.589112	-0.04181	0.508912	-0.27189
UJIR	0.74529	0.200444	0.134623	0.187698	0.048522
LJIR	0.616825	0.31328	0.252757	0.184031	0.437314
CERATO	0.536054	**0.329492**	0.199523	-0.12782	-0.31952
EPI	0.544793	-0.16068	0.155227	-0.08772	0.125343
DSPIN	0.138776	**-0.30468**	-0.19918	0.738834	-0.25709
DSOFT	**0.769377**	0.085868	0.227785	-0.18376	-0.19352
ASOFT	**0.641635**	-0.19394	0.180194	-0.14638	-0.37682
P	0.280804	0.245386	0.699027	0.077348	0.144218
ULL	0.374309	-0.3839	0.269446	0.208206	0.437192
LLL	0.470554	-0.6631	-0.22741	0.060038	-0.23955
LOL	**0.640252**	**-0.64416**	-0.00211	-0.02488	0.052811
DLLA	**0.76511**	-0.1502	-0.28075	-0.2057	0.188191
ALLA	0.426194	**-0.49594**	-0.2207	0.227093	0.231351
PV	0.075488	**0.496354**	-0.56013	0.026975	0.423327
CP	**0.65757**	0.134087	-0.33596	-0.44884	-0.04471
CK	**0.669421**	0.294686	-0.18224	0.15797	-0.07911
Proportion of total variation	0.305802	0.151022	0.087053	0.073795	0.065657

Figure 33c. PCA based on meristic data of S. *altus* and *S. angusticeps*. This plot is between PC1 and PC2. N = 42. The two species separate with minor overlaps both PC1 and PC2. The variables with the highest factor loadings on PC1 and PC2 are set in bold. Tooth counts are ignored.



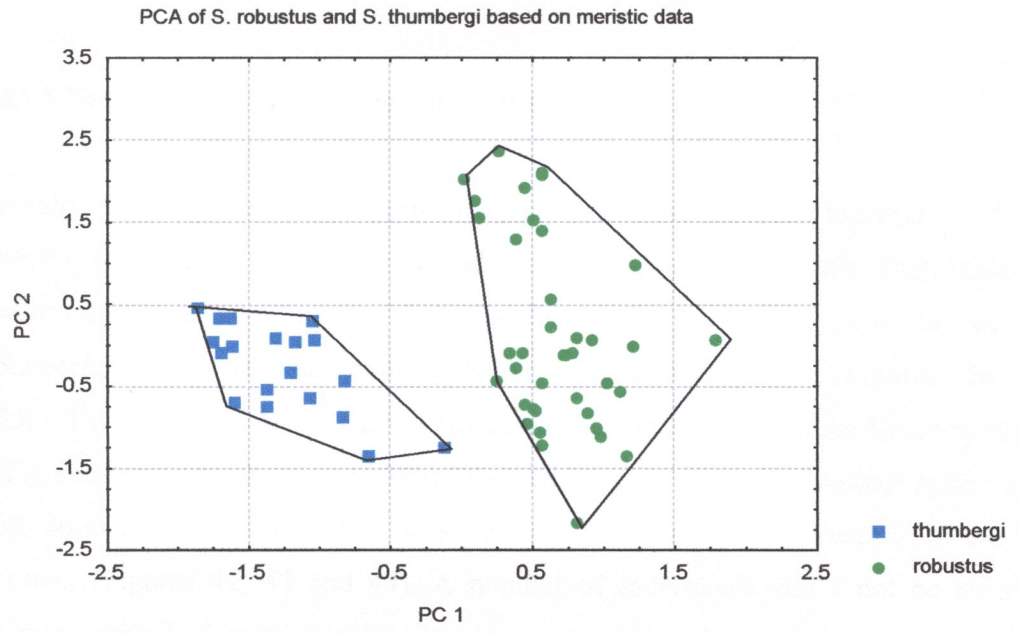
Variable	PC1	PC2	PC3	PC4	PC5
LOGLACRD	0.988956	-0.02724	-0.06069	0.041402	-0.07701
LOGSNL	0.987025	-0.00984	-0.07106	0.026343	0.092105
LOGLJL	0.99207	-0.0099	-0.0576	-0.00401	-0.01941
LOGPPL	0.96154	**0.23083**	0.011935	-0.08014	-0.05841
LOGCHD	0.988225	0.044437	-0.01449	0.022415	-0.02647
LOGED	0.958577	0.076797	**0.265214**	0.028067	0.033972
LOGIOW	0.983064	**0.136866**	-0.07823	0.011346	0.033738
LOGHW	0.986537	**0.107644**	-0.03076	-0.01806	0.078706
LOGHL	0.993627	0.01072	0.04212	-0.0158	-0.05387
LOGSL	0.993494	-0.10376	0.006189	-0.0055	0.010914
LOGBD	0.992879	0.08467	-0.00377	-0.02054	0.012883
LOGDFB	0.981625	**-0.13692**	-0.03461	0.07248	-0.01529
LOGAFB	0.990455	-0.04436	0.00346	-0.07011	-0.03109
LOGPRD	0.988918	0.017427	0.005922	0.059736	-0.00347
LOGPRP	0.98991	-0.03361	0.027672	0.087097	-0.02923
LOGPRV	0.997418	-0.00356	0.015986	0.002545	-0.00122
LOGPRA	0.993978	-0.0632	-0.00695	0.000666	0.024167
LOGCPL	0.940495	**-0.31698**	0.041776	-0.09121	0.01619
LOGCPD	0.992464	0.033717	-0.05251	-0.05155	0.013572
Proportion of total variation	0.969005	0.012538	0.005182	0.002295	0.001772

Figure 34a. PCA based on log-transformed data of *S. robustus* and *S. thumbergi*. This plot is between PC2 and PC3. N = 56. The two species separate completely on PC2. The variables with the highest factor loadings on PC2 and PC3 are set in bold.



Variable	PC1	PC2	PC3	PC4	PC5
LACRD-HL	0.688191	**0.567171**	-0.1475	0.135473	0.045556
SNL-HL	0.599057	0.549135	0.217563	0.234997	-0.19157
LJL-HL	0.708916	0.523255	-0.03006	0.095131	0.041733
PPL-HL	0.561822	-0.43622	0.456535	-0.0539	0.155365
CHD-HL	**0.765156**	0.318264	-0.0424	0.006918	0.063931
ED-HL	-0.6927	-0.2989	0.490441	0.15957	0.033747
IOW-HL	**0.835528**	-0.00733	0.2915	0.18788	-0.13779
IOW-HW	0.25052	-0.26687	-0.57373	0.364267	-0.26395
HW-HL	0.71058	0.130847	0.642871	-0.01578	0.047593
HL-SL	0.249855	**-0.81055**	-0.40438	-0.12773	0.005973
BD-SL	**0.777662**	-0.52044	-0.02668	-0.11928	0.079403
DFB-SL	0.169208	0.369528	-0.26576	0.65294	0.399058
AFB-SL	0.519944	-0.21809	-0.20027	-0.50346	0.3508
PRD-SL	0.268689	**-0.62591**	0.18914	0.32068	0.206102
PRP-SL	-0.17938	**-0.67127**	-0.00393	0.575353	0.097913
PRV-SL	-0.003	**-0.87837**	0.113998	0.076043	-0.04979
PRA-SL	0.499573	-0.28214	0.047486	0.06367	-0.61722
CPL-SL	**-0.871**	0.183885	0.023445	-0.07857	-0.05636
CPD-SL	**0.894297**	-0.14363	-0.15235	-0.17203	-0.02181
CPD-CPL	**0.935895**	-0.19544	-0.1024	-0.05995	0.017162
Proportion of total variation	0.38681	0.213223	0.085421	0.073634	0.044445

Figure 34b. PCA based on percentage data of *S. robustus* and *S. thumbergi*. This plot is between PC1 and PC2. N = 56. The two species separate with minor overlaps on both PC1 and PC2. The variables with the highest factor loadings on PC1 and PC2 are set in bold.



Variable	PC1	PC2	PC3	PC4	PC5
UJT	-0.84	-0.08397	0.25034	0.120038	-0.10103
LJT	-0.91692	0.020411	-0.06759	0.050979	-0.10992
UJIR	-0.71888	-0.32142	0.18855	-0.06621	-0.17289
LJIR	-0.11233	0.192777	-0.44563	0.429051	-0.21015
CERATO	-0.0868	-0.28171	-0.43382	0.409406	0.251455
EPI	0.59276	-0.2853	-0.3853	0.147911	-0.21128
DSPIN	-0.25739	-0.41386	0.328739	0.577229	-0.05528
DSOFT	0.07219	**-0.64383**	-0.35906	-0.06031	0.056905
ASOFT	**0.654211**	**-0.55928**	0.109754	0.083056	-0.07333
P	**0.613187**	0.147483	-0.27795	0.293308	0.335407
ULL	**-0.73823**	-0.28775	-0.12873	-0.08226	0.138618
LLL	-0.33614	**-0.73218**	-0.18858	-0.04058	0.034165
LOL	**-0.67374**	-0.52769	-0.28567	-0.03662	0.11858
DLLA	0.199699	**-0.55738**	0.081444	-0.44941	-0.27328
ALLA	-0.08272	-0.01928	-0.06084	-0.4557	0.728115
PV	0.151006	-0.18955	0.645463	0.385267	0.355744
CP	0.34934	**-0.70562**	0.236212	0.094464	0.191025
CK	**0.744607**	-0.44371	0.109489	-0.16572	-0.23429
Proportion of total variation	0.287112	0.176007	0.088489	0.079385	0.065866

Figure 34c. PCA based on meristic data of *S. robustus* and *S. thumbergi*. This plot is between PC1 and PC2. N = 59. The two species separate completely on PC1. The variables with the highest factor loadings on PC 1 and PC2 are set in bold. Tooth counts are ignored

4.1.5 *Sargochromis* morphological results

Intially all the species of the genus *Sargochromis* were analysed together but the results were taxonomically uninformative and therefore not presented here. Then these species were analysed according to their localities. The first PCA was carried out on *Sargochromis* species from both the Kafue and Zambezi systems (Figures 35a, 35b and 35c). The second PCA was carried out on the *Sargochromis* from the Kafue system alone (Figures 36a, 36b, 37, 38a and 38b) followed by those from the Zambezi system (Figures 39, 40 and 41). The final PCA was carried out on the *Sargochromis* from the Luapula system (Figures 42, 43 and 44). A number of individuals could not be attributed to already-described, species and their working names are given in inverted commas.

4.1.5.1 Zambezi and Kafue *Sargochromis*

The PCA on log-transformed data of 91 individuals of *Sargochromis* from both the Zambezi and Kafue Rivers resulted in the separation of *Sargochromis* 'sp small', '*Pharyngochromis-Sargochromis*-like' species and S. *giardi* from the rest with some overlap, on both PC2 and PC3 as depicted in Figure 35a. *S. codringtonii* separated from the rest with minor overlap on PC2. The variables with the highest loadings on PC2 were body depth (BD) and caudal peduncle length (CPL) while on PC3 the variables with the highest loadings were lachrymal depth (LacrD) and eye depth (ED).

Figure 35b shows the PCA results of the same 91 individuals as above, but based on percentage data. In this analysis two major clusters were formed with the '*codringtonii*-like' individuals separating from the '*carlottae-giardi*-like' individuals on PC1. The variables with the highest loadings on PC1 were inter-orbital width (IOW), body depth (BD), dorsal fin base (DFB), caudal peduncle length (CPL) and caudal peduncle depth (CPD).

PCA results on 66 individual of *Sargochromis* from the Zambezi and Kafue River on meristic data resulted in the separation of *Sargochromis* 'sp small' from the rest, with an

overlap, on PC3 while *S. giardi* completely separates from the rest on PC1. The variables with the highest loadings on PC1 were ceretobranchial gill rakers (CERATO), dorsal fin soft rays (DSOFT), lateral line scales (LOL), scales between the anal fin and lateral line (ALLA) and scales rows on the cheek (CK). On PC2, the variables with highest loadings were epibranchial gill rakers (EPI), dorsal fin spiny rays (DSPIN) and scales between the dorsal fin and lateral line (DLLA). Figure 35c shows these results.

4.1.5.2 Kafue *Sargochromis*

The results from the PCA based on log-transformed data of 45 individuals of *Sargochromis* from the Kafue River drainage show that *Sargochromis 'longsnout kafue'* and *S.* 'sp zambia' separate from the rest on PC2 and from each other on PC3. The variables with the highest loadings on PC2 included snout length (SNL), eye diameter (ED), inter-orbital width (IOW), anal fin base (AFB) and caudal peduncle depth (CPD). The variables with the highest loadings on PC3 were lachrymal depth (LacrD), snout length (SNL), lower jaw length (LJL), eye diameter (ED) and anal fin base (AFB). These results are depicted in Figures 36a and 36b.

Figure 37 shows results from the PCA based on percentage data of the same 45 individuals. *S. codringtonii* separates from the rest with a minor overlap on PC1. *Sargochromis coulteri* clearely separates from the rest on PC2. *Sargochromis* 'sp small' and *Sargochromis* 'sp zambia' separate together from the rest on PC1 and from each other on PC1. The variables with the highest loadings on PC1 included ratios: eye diameter (ED-HL), inter-orbital width (IOW-HL), dorsal fin base (DFB-SL) and caudal peduncle depth (CPD-SL and CPD-CPL). On PC3, the variables with the highest loadings included the ratios; head length (HL-SL), pre-dorsal length (PRD-SL), pre-pelvic length (PRP-SL), pre-ventral length (PRV-SL) and caudal peduncle length (CPL-SL).

PCA results based on meristic data place the 34 individuals of Kafue *Sargochromis* into three clusters on PC2 as depicted in Figure 38a. One of the clusters consists mainly of

Sargochromis 'sp double striped-2' and one individual of *Sargochromis* 'sp double striped-1'. The other two clusters have different species each and overlap with each other. Figure 38b shows four clusters. The first cluster consists of '*carlottae*-like' individuals, which separate from the rest on PC1. On PC2, the remaining three clusters comprising *Sargochromis* 'sp double-striped-2, *Sargochromis* 'sp longsnout-kafue' and *Sargochromis* 'sp double-striped-1 separate from each other with major overlaps. The variables with the highest loadings on PC1 included pectoral fin rays (P), lateral line scales (LLL and LOL) and scale rows on the cheek (CK). On PC2 the variables with the highest loadings included ceratobranchial gill rakers (CERATO), dorsal fin and anal fin soft rays (DSOFT and ASOFT), scales between anal fin and lateral line (ALLA) and scales between the pelvic and ventral fins (PV).

The Mann Whitney U-test was performed on the '*carlottae*-like' individuals to establish the similarity among them. The individuals used in this test were of a similar size range. Tables 10 and 11 show the results of the test carried out on *Sargochromis carlottae* and *Sargochromis* 'sp codringtoni-big' based on percentage and meristic data, respectively. The variables with the highest p-value in Table 10 were LacrD-SL, ED-SL, HW-HL, HL-SL and BD-SL. In Table 11 the variables with the highest p-value were DSPIN, ASOFT, P, LL and CK. Tables 12 and 13 show the results of the Mann Whitnet U-test carried out on *Sargochromis carlottae* and *Sargochromis* 'sp double striped-1' also based on percentage and meristic data, respectively. In Table 12 the variables with the highest p-value were SNL-HL, PPL-HL, HW-HL and CHD-HL while the variables with the highest p-value in Table 13 were DSPIN, LLL, ALLA, PV and CP. These results confirm that *S. carlottae* is morphologically different from *Sargochromis* 'sp codringtoni-big' and *Sargochromis* 'sp double striped-1'.

4.1.5.3 Zambezi *Sargochromis*

PCA results based on log-transformed data of 28 individuals from the Zambezi drainage show that there are four clusters, with a very minor overlap, as depicted in Figure 39. The variables with the highest loading on PC2 were lachrymal depth (LacrD), snout length

(SnL), lower jaw length (LJL), anal fin base (AFB) and caudal peduncle length (CPL). On PC3, the aditional variables with the highest loadings were pre-maxillary pedicel length (PPL) and eye diameter (ED).

The results of the PCA based on percentage data of the same 28 individuals identified three clusters as depicted in Figure 40. *Sargochromis carlottae* and *S. giardi* separate from the '*Pharyngochromis-sargochromis*' on PC1 and from *S. codringtonii* on PC2. The single specimen of *Sargochromis* 'sp longsnout-zambezi' stands out alone. The variables with the highest loadings on PC1 were body depth (BD-SL), caudal peduncle length (CPD-SL), inter-orbital width (IOW-HL), lachrymal depth (LacrD-HL) and anal fin base (AFB-SL). On PC2, the variables with the highest loadings were the ratios; pre-dorsal length (PRD-SL), pre-ventral length (PRV-SL), pre-anal length (PRA-SL), caudal peduncle length (CPL-SL) and caudal peduncle depth (CPD-CPL).

Figure 41 shows the PCA analysis of 32 Kafue *Sargochromis* individuals, based on meristic data. Only two clusters are formed with *S. giardi* separating from the rest on PC1.The variables with the highest loadings on PC1 were scales along the lateral line (ULL and LOL), scales between the lateral line and anal fin (ALLA), scales around the caudal peduncle (CP) and scales rows on cheek (CK).

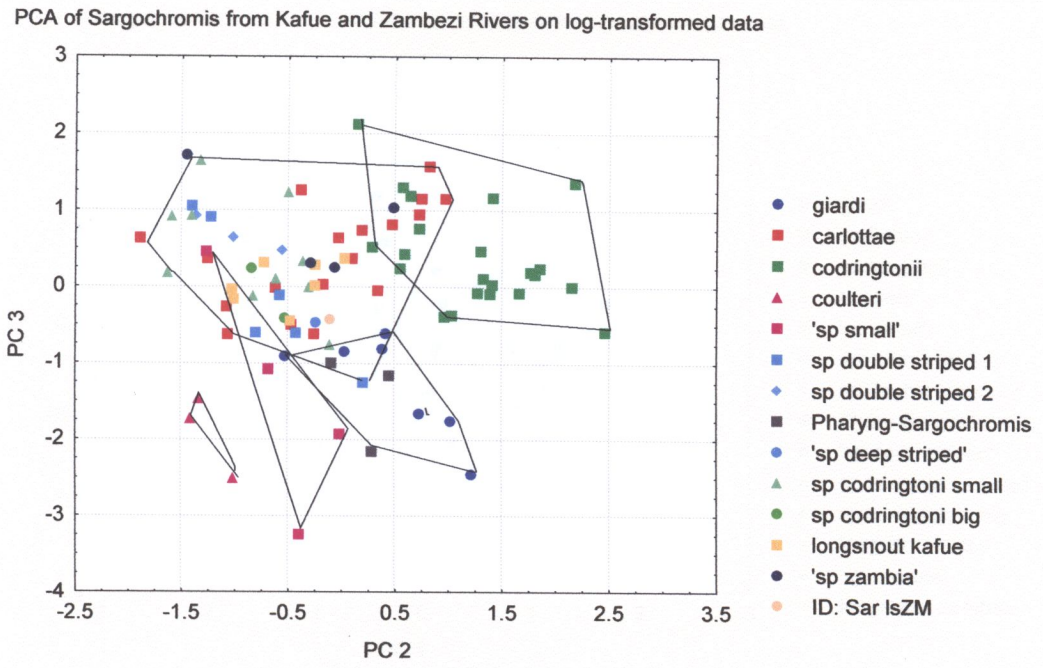
4.1.5.4 Luapula *Sargochromis*

The results of the PCA based log-transformed data of 22 individuals from the Luapula River drainage resulted in one cluster with no morphological difference between the individuals. This result is depicted in Figure 42.

Figure 43 shows the PCA analysis of the same 22 individuals, based on percentage data. The individuals separate into two groups, with an overlap on both PC1 and PC2. The variables with the highest loadings on PC1 were the ratios; cheek depth (CHD-HL), head length (HL-SL), pre-ventral length (PRV-SL) and caudal peduncle depth (CPD-SL and CPD-CPL). On PC2, the variables with the highest loadings were the ratios; snout length

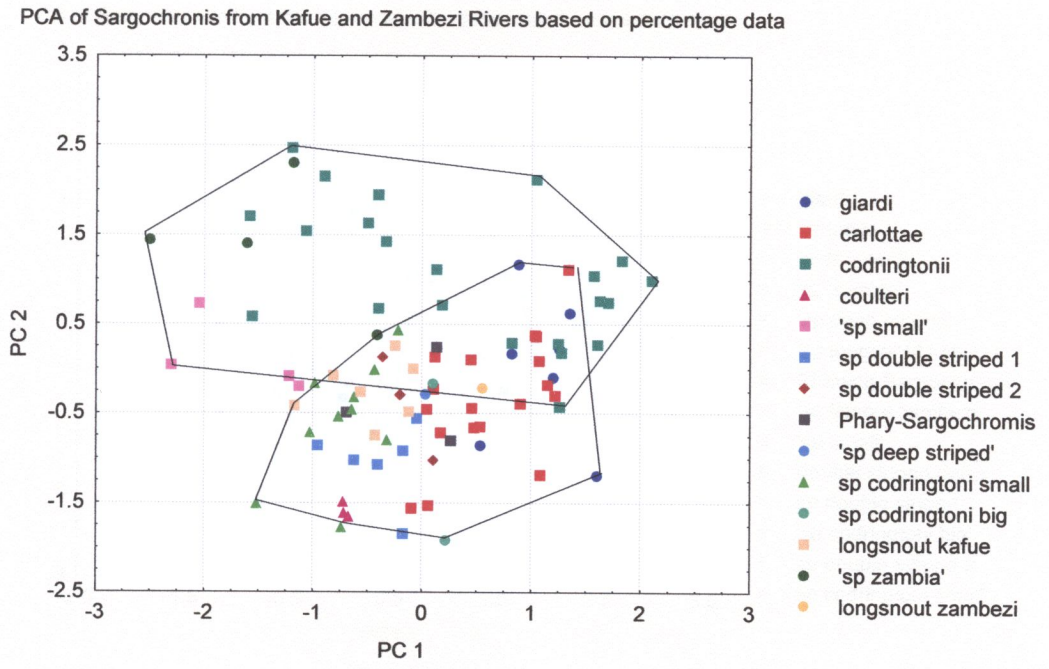
(SNL-HL), pre-maxillary pedicel length (PPL-HL), inter-orbital width (IOW-HL and IOW-HW) and dorsal fin base (DFB-SL).

The results of the analysis based on meristic data of 20 individuals places them into two groups as seen in Figure 44. The separation is on PC2 with a minor overlap. The variables with the highest loadings on PC2 were dorsal fin soft rays (DSOFT), lateral line scales (ULL), scales between lateral line and both dorsal and anal fins (DLLA and ALLA), and scales around the caudal peduncle (CP).



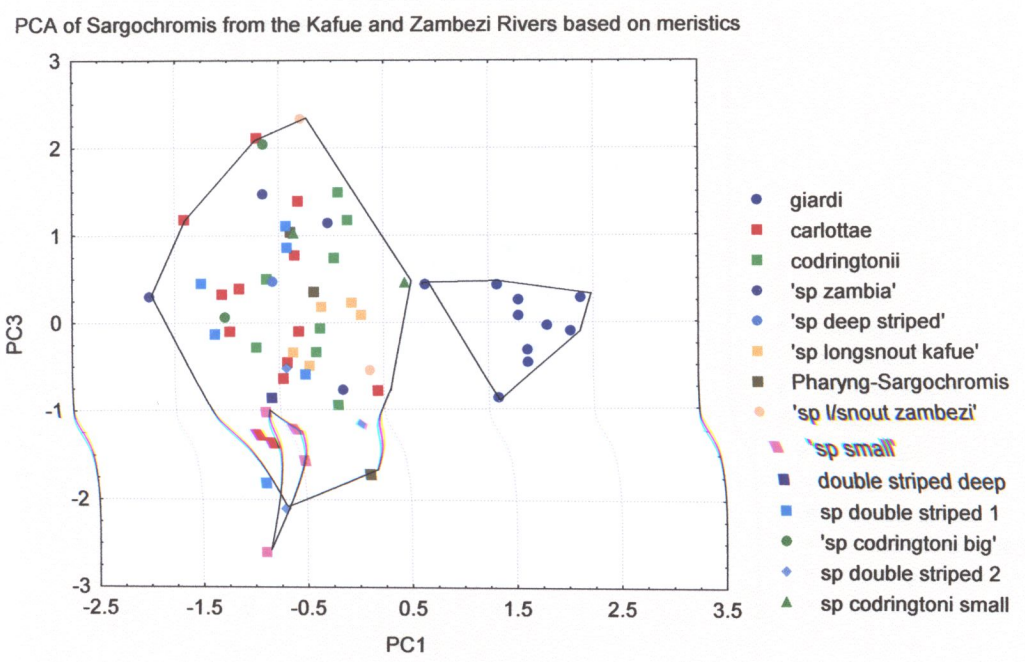
Variables	PC1	PC2	PC3	PC4	PC5
LOGLACRD	0.94941	0.055169	**-0.17035**	-0.229	0.020264
LOGSNL	0.965374	0.022186	-0.0546	-0.18341	-0.13299
LOGLJL	0.972088	0.007983	0.06922	0.019854	0.126153
LOGPPL	0.970774	-0.02539	0.054696	-0.10646	0.148247
LOGCHD	0.978255	0.059424	0.031068	-0.08908	0.046853
LOGED	0.919495	0.013774	**0.375644**	-0.06286	-0.02792
LOGIOW	0.979726	0.045067	-0.07564	0.100066	0.023274
LOGHW	0.967208	0.020669	-0.08963	0.010287	0.095709
LOGHL	0.994813	0.018948	0.000998	-0.01564	-0.04521
LOGSL	0.994901	-0.0642	-0.02451	0.026773	-0.02348
LOGBD	0.981178	**0.100979**	-0.01064	0.107524	-0.00025
LOGDFB	0.990669	-0.01154	-0.04807	0.060238	-0.0033
LOGAFB	0.967793	0.03583	-0.00013	0.172917	-0.02217
LOGPRD	0.993092	0.021758	-0.00391	-0.01619	-0.03954
LOGPRP	0.992311	0.014197	0.021487	0.04509	-0.04094
LOGPRV	0.987386	0.042121	0.031643	0.00204	-0.08016
LOGPRA	0.994535	0.013611	-0.02346	0.020504	-0.01814
LOGCPL	0.879045	**-0.4741**	-0.02139	0.01323	-0.01084
LOGCPD	0.972567	0.058383	-0.04722	0.112269	-0.01339
Proportion of total variation	0.943804	0.013544	0.010703	0.009546	0.004319

Figure 35a. PCA based on log-transformed data, of the Zambezi and Kafue River *Sargochromis*. This plot is between PC2 and PC3. N = 91.Five clusters result from this PCA. The variables with the highest factor loadings on PC2 and PC3 are set in bold.



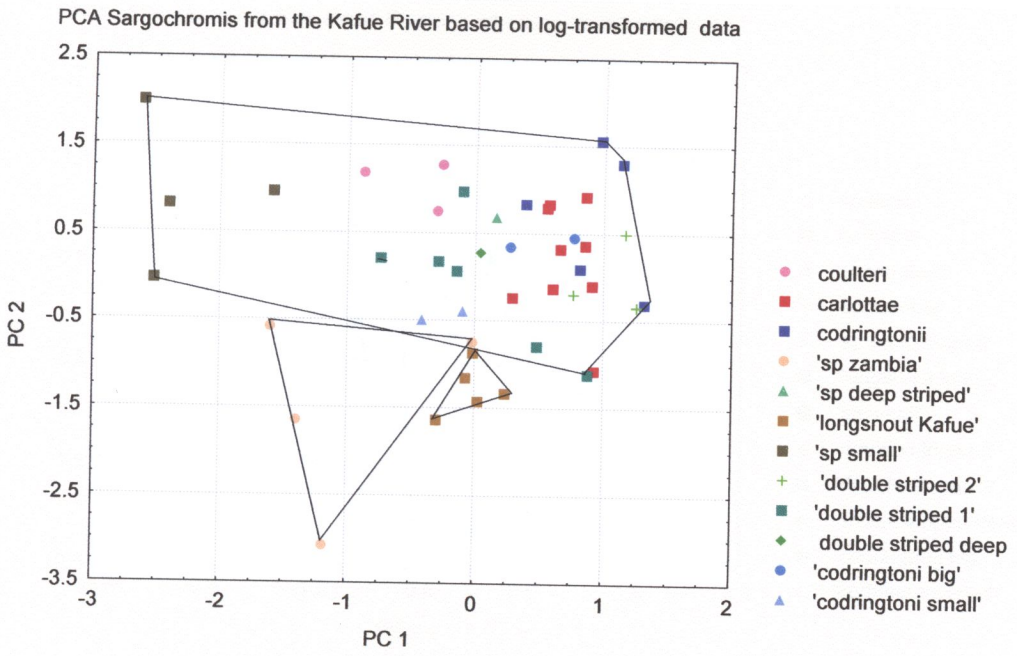
Variables	PC1	PC2	PC3	PC4	PC5
LACRD-HL	0.431503	-0.36179	-0.64523	0.209681	-0.11362
SNL-HL	0.232226	-0.21111	-0.68238	0.305023	-0.2565
LJL-HL	0.38759	-0.27177	0.138719	-0.59927	-0.39846
PPL-HL	0.065439	-0.39055	-0.2035	-0.53209	-0.48934
CHD-HL	0.632544	-0.33732	-0.2876	-0.21277	-0.13334
ED-HL	-0.64437	0.374113	0.101881	-0.38128	-0.067
IOW-HL	**0.769581**	-0.37634	0.280651	0.139333	-0.06285
IOW-HW	0.371575	0.023287	0.503472	0.451685	-0.60527
HW-HL	0.441976	-0.48255	-0.22635	-0.31368	0.568917
HL-SL	-0.01188	**0.861037**	-0.23787	0.014156	0.018154
BD-SL	**0.798251**	0.294797	0.134787	-0.13704	0.132374
DFB-SL	**0.737213**	-0.19043	0.117076	0.017457	0.232953
AFB-SL	0.647255	-0.00908	0.372891	-0.12774	0.141787
PRD-SL	0.246106	**0.632372**	-0.42427	-0.09582	-0.12161
PRP-SL	0.215492	**0.766675**	0.110865	-0.22313	-0.11964
PRV-SL	-0.07391	**0.839971**	-0.12852	-0.06014	-0.00708
PRA-SL	0.605838	0.43916	-0.12965	0.020907	-0.08031
CPL-SL	**-0.74335**	-0.42426	0.134973	-0.02991	-0.07846
CPD-SL	0.681942	0.116258	0.221345	0.039937	-0.02659
CPD-CPL	**0.8471**	0.338109	0.002197	0.050432	0.062257
Proportion of total variation	0.297504	0.204459	0.096977	0.069769	0.066587

Figure 35b. PCA based on percentage data of of the Zambezi and Kafue River *Sargochromis*. This plot is between PC1 and PC2. N = 91.The different species separate into two major clusters ('*codringtonii*-like' and '*carlottae-giardi*-like') on PC2. The variables with the highest factor loadings on PC1 and PC2 are set in bold.



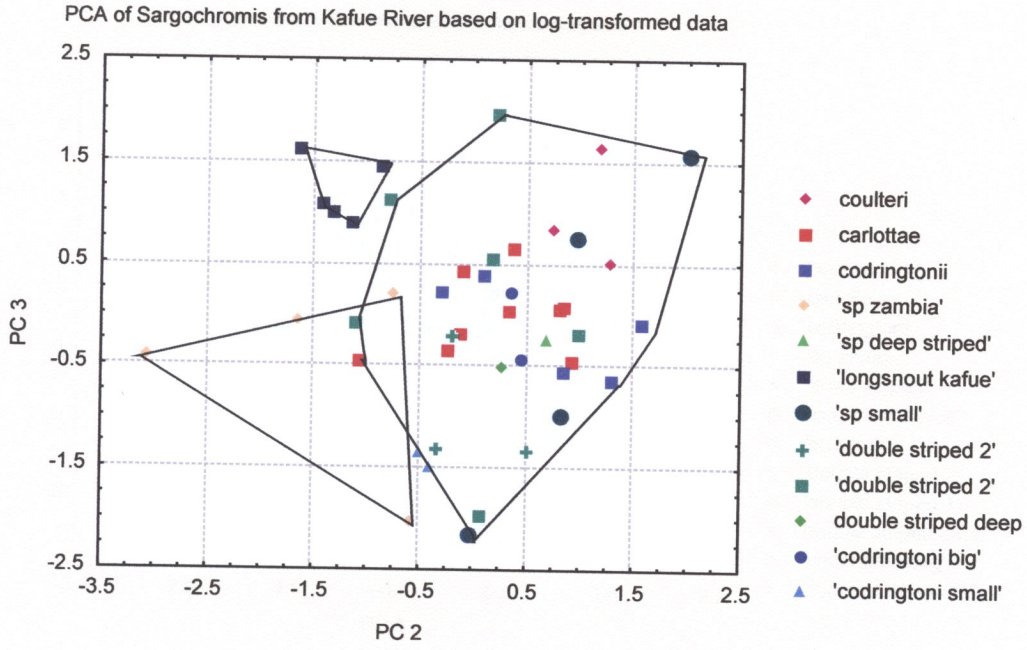
Variables	PC1	PC2	PC3	PC4	PC5
UJT	-0.2507	0.652037	-0.02987	-0.40927	-0.15297
LJT	-0.41685	0.429282	-0.16502	-0.42011	-0.02168
UJIR	-0.49435	0.503026	0.380648	0.15576	0.0083
LJIR	-0.42052	0.679634	0.193071	0.341676	0.039256
CERATO	**-0.66798**	0.098895	0.060655	0.171624	0.288028
EPI	0.048977	0.110687	**0.567406**	0.046833	-0.48619
DSPIN	0.161355	0.000111	**0.553013**	-0.03512	0.675056
DSOFT	**0.634166**	0.122396	0.10063	0.11788	0.045993
ASOFT	0.514274	0.27306	-0.19367	0.285022	0.417441
P	0.429734	0.527974	0.119282	-0.02077	-0.11065
ULL	0.530305	-0.16603	0.058435	-0.17792	-0.23766
LLL	0.487841	0.161165	0.4125	-0.50334	0.154539
LOL	**0.807822**	-0.00651	0.178341	-0.08989	-0.144
DLLA	0.214576	0.547632	**-0.43796**	0.270565	-0.23423
ALLA	**0.642361**	0.146668	-0.37448	0.146621	0.209235
PV	0.307216	0.096142	0.309306	0.611316	-0.20919
CP	0.038789	0.560835	-0.18051	-0.1996	0.148085
CK	**0.733742**	0.332866	-0.03909	-0.09708	0.034826
Proportion of total variation	0.236028	0.140238	0.087267	0.079122	0.069766

Figure 35c. PCA based on meristic data of of the Zambezi and Kafue River *Sargochromis*. This plot is between PC1and PC3. N = 66. *S. giardi* clearly separates from the rest on PC1 while *Sargochromis* 'sp small' separates from the rest, with an overlap, on PC3. The variables with the highest factor loadings on PC1 and PC3 are set in bold.



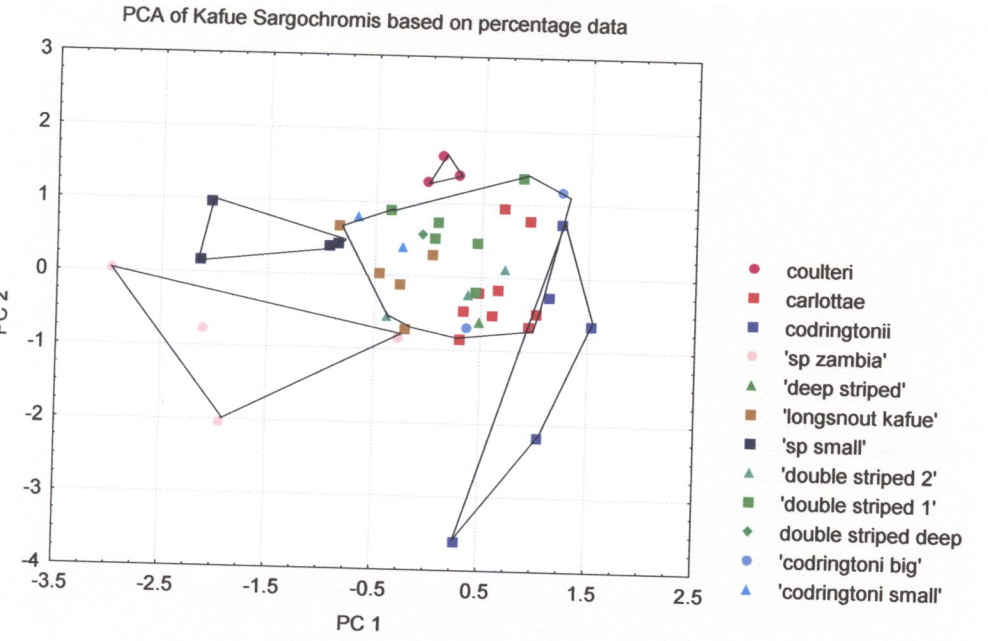
Variable	PC1	PC2	PC3	PC4	PC5
LOGLACRD	0.967511	-0.0394	0.228319	0.027844	-0.01154
LOGSNL	0.965707	**-0.1577**	0.139973	-0.01631	-0.10506
LOGLJL	0.974005	-0.03707	0.078096	-0.03176	0.138792
LOGPPL	0.983312	-0.05716	0.066393	0.007231	0.012806
LOGCHD	0.989951	-0.07343	0.022484	0.017083	-0.00237
LOGED	0.933655	**-0.27457**	-0.21621	-0.0188	0.013982
LOGIOW	0.965707	**0.137412**	-0.03439	-0.11831	0.112927
LOGHW	0.974879	0.084797	0.002431	0.056247	0.10398
LOGHL	0.993854	-0.05059	-0.01113	-0.04716	-0.00665
LOGSL	0.99687	0.039098	-0.01569	0.039651	-0.02056
LOGBD	0.981082	**0.109243**	-0.03127	-0.08179	-0.06368
LOGDFB	0.988669	**0.101355**	-0.00425	0.022919	-0.02909
LOGAFB	0.965207	**0.160285**	-0.0944	0.038923	-0.13286
LOGPRD	0.993347	-0.01648	0.007766	-0.03083	-0.05367
LOGPRP	0.993894	-0.01699	-0.03722	-0.0431	-0.00108
LOGPRV	0.988808	-0.06431	-0.04855	-0.07887	0.006615
LOGPRA	0.994492	0.03789	-0.00365	0.001875	0.006352
LOGCPL	0.954447	-0.00932	-0.05246	0.279468	0.037933
LOGCPD	0.985613	**0.115323**	-0.00334	-0.01735	-0.00443
Proportion of total variation	0.957673	0.011022	0.007765	0.006309	0.004306

Figure 36a. PCA based on log-transformed data, of the Kafue River *Sargochromis*. This plot is between PC1 and PC2. N = 45. *Sargochromis* 'sp zambia' and *Sargochromis* 'sp longsnout kafue', separate with a minor overlap from the rest on PC2. The variables with the highest factor loadings on PC2 are set in bold.



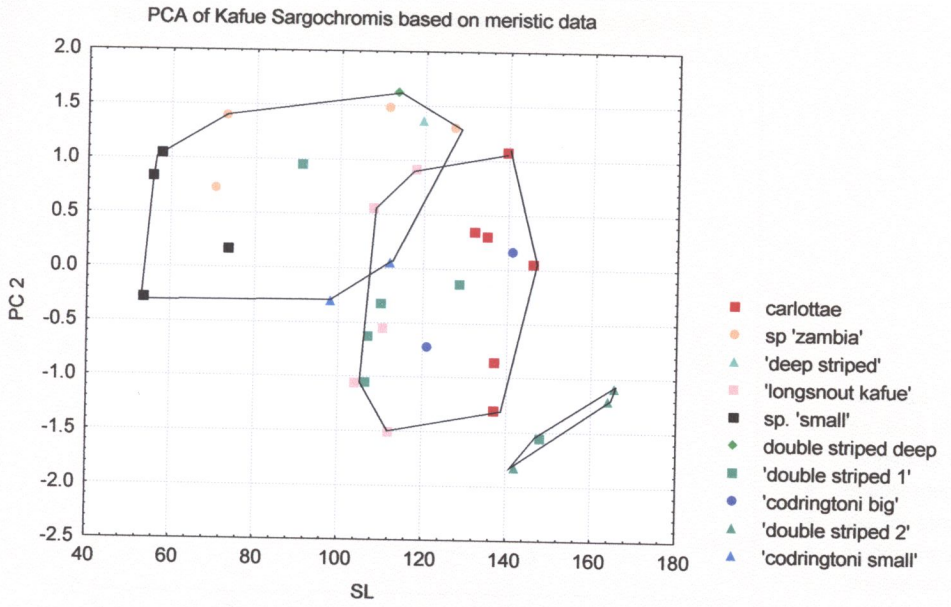
Variable	PC1	PC2	PC3	PC4	PC5
LOGLACRD	0.967511	-0.0394	**0.228319**	0.027844	-0.01154
LOGSNL	0.965707	**-0.1577**	**0.139973**	-0.01631	-0.10506
LOGLJL	0.974005	-0.03707	**0.078096**	-0.03176	0.138792
LOGPPL	0.983312	-0.05716	0.066393	0.007231	0.012806
LOGCHD	0.989951	-0.07343	0.022484	0.017083	-0.00237
LOGED	0.933655	**-0.27457**	**-0.21621**	-0.0188	0.013982
LOGIOW	0.965707	**0.137412**	-0.03439	-0.11831	0.112927
LOGHW	0.974879	0.084797	0.002431	0.056247	0.10398
LOGHL	0.993854	-0.05059	-0.01113	-0.04716	-0.00665
LOGSL	0.99687	0.039098	-0.01569	0.039651	-0.02056
LOGBD	0.981082	**0.109243**	-0.03127	-0.08179	-0.06368
LOGDFB	0.988669	**0.101355**	-0.00425	0.022919	-0.02909
LOGAFB	0.965207	**0.160285**	**-0.0944**	0.038923	-0.13286
LOGPRD	0.993347	-0.01648	0.007766	-0.03083	-0.05367
LOGPRP	0.993894	-0.01699	-0.03722	-0.0431	-0.00108
LOGPRV	0.988808	-0.06431	-0.04855	-0.07887	0.006615
LOGPRA	0.994492	0.03789	-0.00365	0.001875	0.006352
LOGCPL	0.954447	-0.00932	-0.05246	0.279468	0.037933
LOGCPD	0.985613	**0.115323**	-0.00334	-0.01735	-0.00443
Proportion of total variation	0.957673	0.011022	0.007765	0.006309	0.004306

Figure 36b. PCA based on log-transformed data, of the Kafue River *Sargochromis*. This plot is between PC1 and PC2. N = 45. *Sargochromis* 'sp zambia' and *Sargochromis* 'sp longsnout kafue', separate from the rest on PC2 and from each other on PC3. The variables with the highest factor loadings on PC2 and PC3 are set in bold.



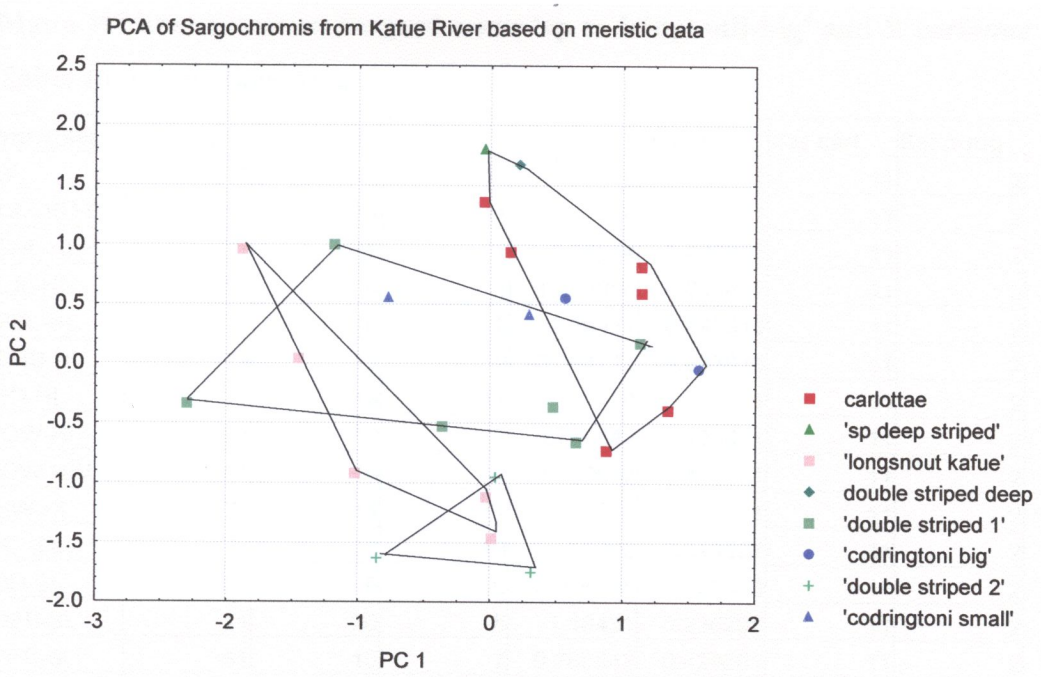
riable	PC1	PC2	PC3	PC4	PC5
CRD-HL	0.597994	0.2341	-0.65085	0.043734	0.008101
L-HL	0.184268	0.066524	-0.81974	-0.00316	0.33597
-HL	0.535242	0.077742	-0.33724	0.294204	-0.3747
-HL	0.448726	0.388432	-0.34053	0.467134	0.124073
D-HL	0.704472	0.009926	-0.35633	-0.06867	-0.12017
-HL	**-0.78448**	0.187316	0.196398	0.243	0.130522
V-HL	**0.722291**	-0.19027	0.320696	0.310944	-0.11028
V-HW	0.331031	-0.46854	0.250217	0.648134	0.331877
-HL	0.52514	0.355431	0.099057	-0.41425	-0.51808
-SL	-0.57456	**-0.66091**	-0.20954	-0.07664	-0.01051
SL	0.578404	-0.59603	0.085064	-0.24037	-0.0066
3-SL	**0.735483**	0.040378	0.242436	-0.13921	0.07804
3-SL	**0.649384**	-0.04613	0.306906	-0.26709	0.324164
D-SL	-0.28812	**-0.63781**	-0.3986	-0.35766	0.206384
P-SL	-0.45266	**-0.75062**	-0.09101	0.148483	-0.202
V-SL	-0.53656	**-0.64523**	-0.14206	0.001693	-0.09948
A-SL	0.036068	-0.50392	-0.08097	0.26625	-0.63954
-SL	-0.52882	**0.638412**	0.027409	0.088693	-0.1226
-SL	**0.782841**	-0.31093	0.129919	0.030853	0.047785
-CPL	**0.716359**	**-0.59693**	0.067587	-0.03658	0.101049
ortion of total variation	0.325411	0.196514	0.104315	0.07221	0.066604

re 37. PCA based on percentage data, of the Kafue River *Sargochromis*. This plot is between PC1 and N = 45. *Sargochromis coulteri, Sargochromis* 'sp *zambia' and Sargochromis* 'sp small' clearly ate from the rest. *Sargochromis codringtonii* separates from the rest as shown in the figure, with overlap. The variables with the highest factor loadings on PC1 and PC2 are set in bold.



Variable	PC1	PC2	PC3	PC4	PC5
UJT	0.642148	-0.55033	-0.122	-0.04781	-0.12242
LJT	0.472538	-0.0872	0.040297	0.535253	-0.07084
UJIR	0.45516	0.224626	0.119196	-0.48358	-0.61228
LJIR	0.562736	0.389024	0.002286	-0.25651	-0.56059
CERATO	-0.01931	**0.602872**	0.241902	-0.18106	0.296221
EPI	0.272478	-0.15577	0.23587	-0.56504	0.189831
DSPIN	0.116565	0.185634	0.455329	-0.18536	0.084794
DSOFT	0.463705	**0.365**	0.255767	0.475486	0.094208
ASOFT	-0.11099	**0.567898**	-0.35578	-0.19872	0.022763
P	0.632266	0.028578	0.2749	0.428435	0.186837
ULL	0.10638	-0.20193	-0.47727	-0.49416	0.394981
LLL	0.663688	-0.30638	0.320741	-0.04849	0.320172
LOL	0.576795	0.141494	-0.41964	-0.26355	0.560011
DLLA	0.232278	0.00949	-0.77225	0.189981	-0.21871
ALLA	0.047141	**0.316238**	-0.51552	0.419782	0.102297
PV	0.319856	**0.750632**	-0.0522	-0.01766	0.113911
CK	0.463107	-0.28591	-0.30337	-0.04455	-0.1661
Proportion of total variation	0.178367	0.134761	0.122553	0.114976	0.091495

Figure 38a. PCA based on meristic data, of the Kafue River *Sargochromis*. This plot is between SL and PC2. N = 34. Three clusters are evident in this plot. The variables with the highest factor loadings on PC2 are set in bold. Tooth counts are ignored.



Variable	PC1	PC2	PC3	PC4	PC5
UJT	0.642148	-0.55033	-0.122	-0.04781	-0.12242
LJT	0.472538	-0.0872	0.040297	0.535253	-0.07084
UJIR	0.45516	0.224626	0.119196	-0.48358	-0.61228
LJIR	0.562736	0.389024	0.002286	-0.25651	-0.56059
CERATO	-0.01931	**0.602872**	0.241902	-0.18106	0.296221
EPI	0.272478	-0.15577	0.23587	-0.56504	0.189831
DSPIN	0.116565	0.185634	0.455329	-0.18536	0.084794
DSOFT	**0.463705**	**0.365**	0.255767	0.475486	0.094208
ASOFT	-0.11099	**0.567898**	-0.35578	-0.19872	0.022763
P	**0.632266**	0.028578	0.2749	0.428435	0.186837
ULL	0.10638	-0.20193	-0.47727	-0.49416	0.394981
LLL	**0.663688**	-0.30638	0.320741	-0.04849	0.320172
LOL	**0.576795**	0.141494	-0.41964	-0.26355	0.560011
DLLA	0.232278	0.00949	-0.77225	0.189981	-0.21871
ALLA	0.047141	**0.316238**	-0.51552	0.419782	0.102297
PV	0.319856	**0.750632**	-0.0522	-0.01766	0.113911
CK	**0.463107**	-0.28591	-0.30337	-0.04455	-0.1661
Proportion of total variation	0.178367	0.134761	0.122553	0.114976	0.091495

Figure 38b. PCA based on meristic data, of the Kafue River *Sargochromis.* This plot is between PC1 and PC2. N = 26. Four clusters are evident in this plot. *Sargochromis* 'sp double striped 1' is likely to comprise morphologically different specimens. The variables with the highest factor loadings on PC1 and PC2 are set in bold. Tooth counts are ignored

Mann Whitnet U-test on *Sargochromis* 'sp codringtonii-big' and *S. carlottae*

Table 10. Percentage data

Variables	Sar carl	Sar cobg	U	Z	p-level	Sar carl	Sar cobg
SL	77	14	11	0	1	11	2
LACRD-HL	77	14	11	0	**1**	11	2
SNL-HL	73	18	7	-0.78954	0.429801	11	2
LJL-HL	87	4	1	1.973855	0.048407	11	2
PPL-HL	75	16	9	-0.39477	**0.693015**	11	2
CHD-HL	79	12	9	0.394771	**0.693015**	11	2
ED-HL	77	14	11	0	1	11	2
IOW-HL	81	10	7	0.789542	0.429801	11	2
IOW-HW	81	10	7	0.789542	0.429801	11	2
HW-HL	77	14	11	0	1	11	2
HL-SL	78	13	10	0.197386	**0.843527**	11	2
BD-SL	82	9	6	0.986928	0.323686	11	2
DFB-SL	71	20	5	-1.18431	0.236298	11	2
AFB-SL	81	10	7	0.789542	0.429801	11	2
PRV-SL	81	10	7	0.789542	0.429801	11	2
PRA-SL	80	11	8	0.592157	**0.55375**	11	2
CPD-SL	83	8	5	1.184313	0.236298	11	2

Table 11. Meristics

Variables	Sar carl	Sar cobg	U	Z	p-level	Sar carl	Sar cobg
SL	53.5	12.5	8.5	-0.11785	0.906186	9	2
HL	54	12	9	0	1	9	2
UJT	57	9	6	0.707107	0.479505	9	2
UJIR	49	17	4	-1.17851	0.238602	9	2
LJIR	55.5	10.5	7.5	0.353553	0.723676	9	2
EPI	56	10	7	0.471405	0.637355	9	2
DSPIN	53	13	8	-0.2357	**0.813665**	9	2
DSOFT	59	7	4	1.178511	0.238602	9	2
ASOFT	55.5	10.5	7.5	0.353553	**0.723676**	9	2
P	52.5	13.5	7.5	-0.35355	**0.723676**	9	2
ULL	51	15	6	-0.70711	0.479505	9	2
LLL	55.5	10.5	7.5	0.353553	**0.723676**	9	2
LOL	52	14	7	-0.4714	0.637355	9	2
ALLA	57.5	8.5	5.5	0.824958	0.409402	9	2
PV	55	11	8	0.235702	0.813665	9	2
CK	53.5	12.5	8.5	-0.11785	**0.906186**	9	2

For both Tables 10 and 11 the p-value for standard length is >0.5 and therefore, the results are valid. The characters with the highest p-values i.e. contributing the most to the observed differences between the two populations are set in bold. Tooth counts are ignored.

Mann Whitnet U-test on *Sargochromis* 'sp double striped-1' and *S. carlottae*

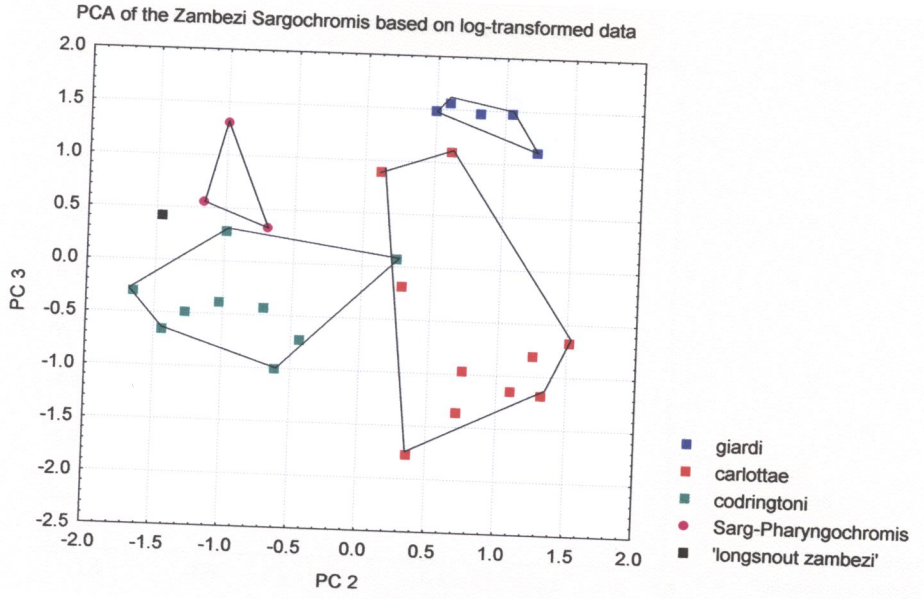
Table 12. Percentage data

Variables	Sar carl	Sar dst1	U	Z	p-level	Sar carl	Sar dst1
SL	30	25	10	0.522233	0.601512	5	5
LACRD-HL	21	34	6	-1.35781	0.174535	5	5
SNL-HL	26	29	11	-0.31334	**0.754025**	5	5
LJL-HL	36	19	4	1.775592	0.07581	5	5
PPL-HL	26	29	11	-0.31334	**0.754025**	5	5
CHD-HL	33	22	7	1.148913	0.250601	5	5
ED-HL	21	34	6	-1.35781	0.174535	5	5
IOW-HL	34	21	6	1.357806	0.174535	5	5
IOW-HW	35	20	5	1.566699	0.117195	5	5
HW-HL	27	28	12	-0.10445	**0.916816**	5	5
DFB-SL	34	21	6	1.357806	0.174535	5	5
AFB-SL	36	19	4	1.775592	0.07581	5	5
PRD-SL	34	21	6	1.357806	0.174535	5	5
PRP-SL	39	16	1	2.402272	0.016299	5	5
PRA-SL	36	19	4	1.775592	0.07581	5	5
CPL-SL	16	39	1	-2.40227	0.016299	5	5
CPD-SL	38	17	2	2.193378	0.028287	5	5
CPD-CPL	40	15	0	2.611165	0.009028	5	5

Table 13. Meristics

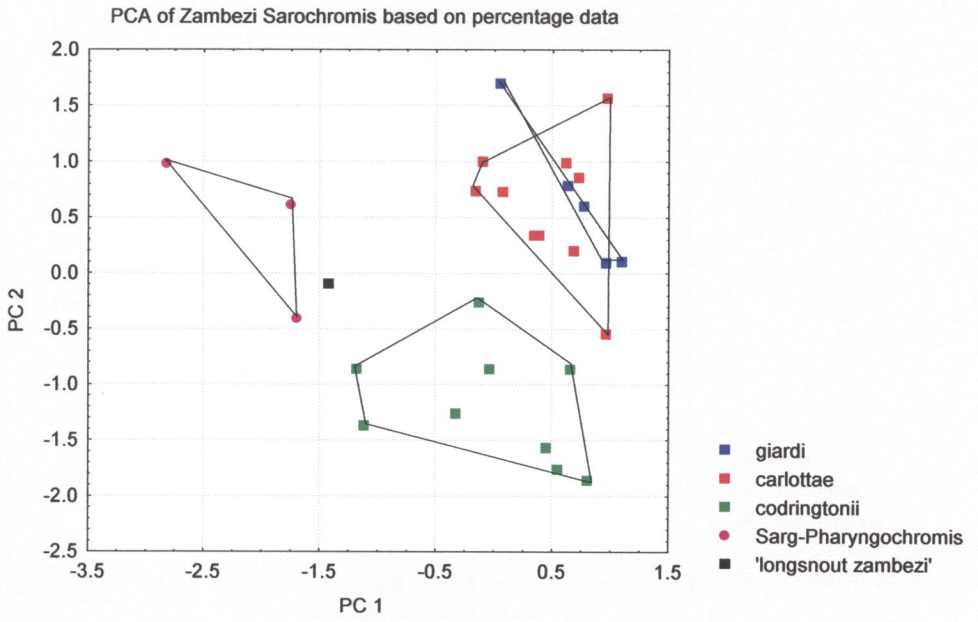
Variables	Sar carl	Sar dst1	U	Z	p-level	Sar carl	Sar dst1
SL	29	26	11	0.31334	0.754025	5	5
HL	31	24	9	0.731126	0.464707	5	5
UJT	29.5	25.5	10.5	0.417786	0.676106	5	5
LJIR	24	31	9	-0.73113	0.464707	5	5
CERATO	30	25	10	0.522233	0.601512	5	5
DSPIN	27.5	27.5	12.5	0	**1**	5	5
DSOFT	38	17	2	2.193378	0.028287	5	5
ASOFT	37.5	17.5	2.5	2.088932	0.036722	5	5
P	26.5	28.5	11.5	-0.20889	0.834533	5	5
LLL	27.5	27.5	12.5	0	**1**	5	5
LOL	26.5	28.5	11.5	-0.20889	0.834533	5	5
DLLA	25	30	10	-0.52223	0.601512	5	5
ALLA	27.5	27.5	12.5	0	**1**	5	5
PV	27.5	27.5	12.5	0	**1**	5	5
CP	27.5	27.5	12.5	0	**1**	5	5

For both Tables 12 and 13 the p-value for standard length is >0.5 and therefore, the results are valid. The characters with the highest p-values i.e. contributing the most to the observed differences between the two populations are set in bold.



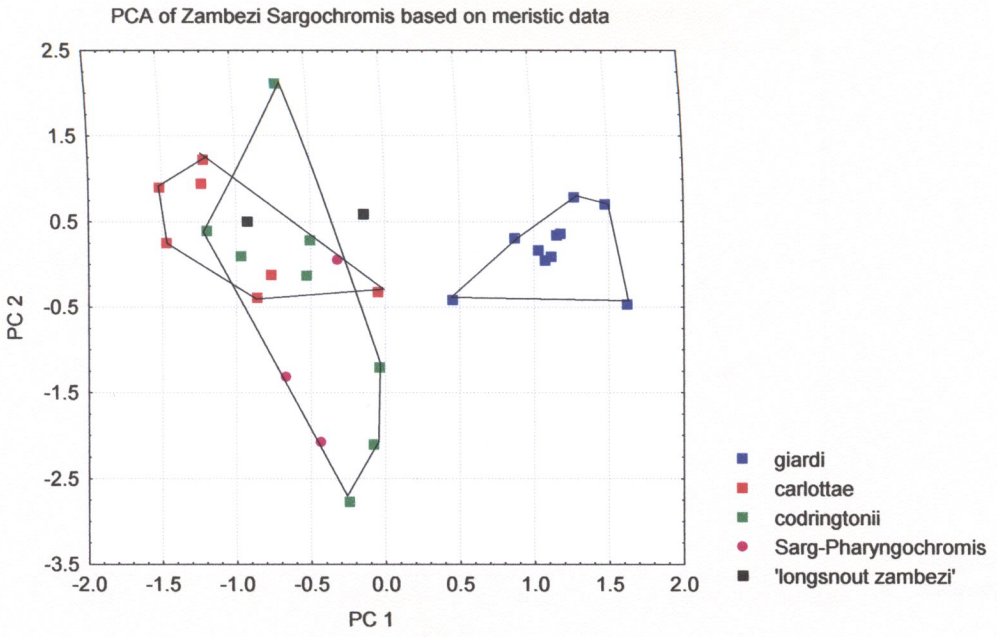
Variable	PC1	PC2	PC3	PC4	PC5
LOGLACRD	0.816207	**-0.51702**	**0.157483**	-0.18046	0.000802
LOGSNL	0.905338	**-0.38961**	0.09281	-0.0135	-0.03657
LOGLJL	0.919067	**0.176355**	**-0.22407**	-0.13921	0.181855
LOGPPL	0.896863	-0.03713	**-0.32527**	-0.23782	0.054194
LOGCHD	0.947862	-0.13741	-0.05299	-0.07014	0.143145
LOGED	0.833333	-0.00348	**-0.49108**	0.013534	-0.23459
LOGIOW	0.95674	0.162047	0.11966	0.090193	0.041014
LOGHW	0.965888	-0.04421	0.007646	0.001497	-0.03826
LOGHL	0.991076	-0.00539	0.054844	0.052902	-0.02234
LOGSL	0.993411	0.01933	0.083137	0.023917	-0.00477
LOGBD	0.980123	0.140744	-0.00565	0.088668	0.011451
LOGDFB	0.983638	0.03038	0.069038	0.106576	-0.03873
LOGAFB	0.960964	**0.176548**	0.046747	0.009334	0.098793
LOGPRD	0.983522	-0.05842	-0.03272	0.075995	0.014649
LOGPRP	0.990034	0.032796	0.022417	0.046426	0.02937
LOGPRV	0.975667	-0.02221	0.069484	0.143063	-0.07262
LOGPRA	0.987006	-0.02492	0.035692	0.099372	-0.00425
LOGCPL	0.781465	**0.334499**	**0.29472**	-0.40638	-0.15224
LOGCPD	0.971392	0.128944	0.056723	0.154958	-0.0169
Proportion of total variation	0.885509	0.036058	0.029601	0.019581	0.008303

Figure 39. PCA based on log-transformed data, of the Zambezi River *Sargochromis*. This plot is between PC2 and PC3. N = 28. All the species separate from each other. *Sargochromis* 'sp longsnout-zambezi' and '*Sargochromis-pharyngochromis*' are morphologically similar. The variables with the highest factor loadings on PC2 and PC3 are set in bold.



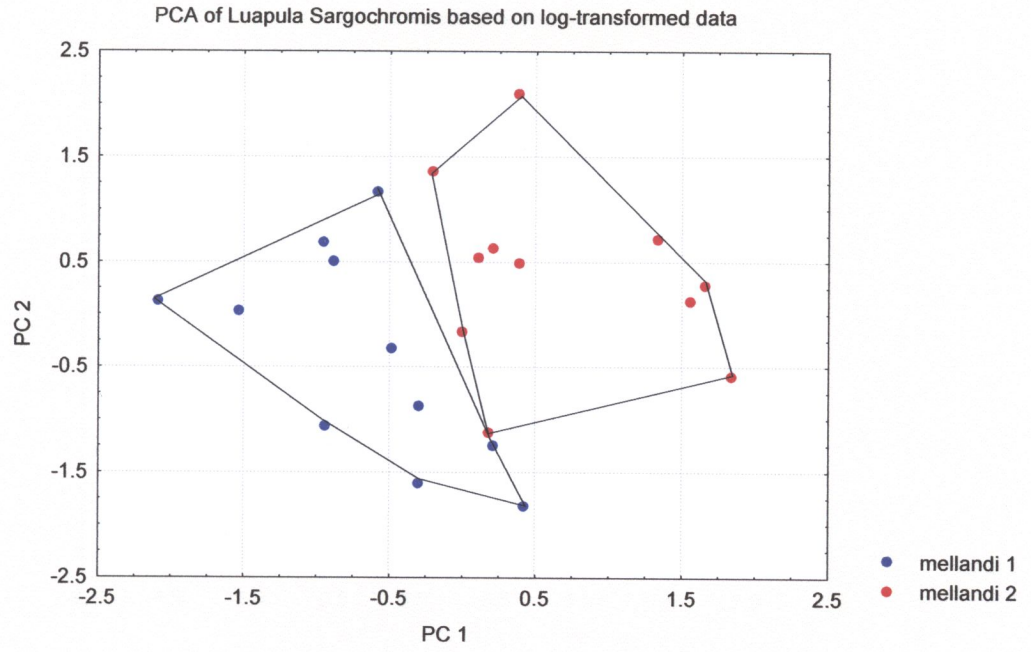
Variable	PC1	PC2	PC3	PC4	PC5
LACRD-HL	**-0.64564**	-0.22641	-0.4648	0.234421	-0.38521
SNL-HL	-0.47247	-0.45032	-0.45174	0.206258	-0.2969
LJL-HL	0.305848	0.360349	-0.36117	-0.70267	-0.02819
PPL-HL	-0.36092	0.082178	-0.23868	-0.79034	0.020554
CHD-HL	0.088716	0.0362	-0.7638	-0.15918	-0.27823
ED-HL	-0.41413	-0.17739	0.185253	-0.61828	0.33599
IOW-HL	**0.780061**	0.367112	-0.13021	0.22729	-0.11443
IOW-HW	0.594012	0.226876	0.424381	0.198929	-0.48117
HW-HL	0.087401	0.089265	-0.73066	0.005011	0.536084
HL-SL	-0.16514	-0.45641	0.604105	-0.21448	-0.21959
BD-SL	**0.866917**	0.024985	-0.02139	-0.29059	-0.00199
DFB-SL	0.557209	-0.31615	-0.02788	0.270334	0.453278
AFB-SL	**0.641527**	0.362891	0.022976	-0.30465	-0.20366
PRD-SL	-0.01174	**-0.68812**	0.074601	-0.4807	-0.22029
PRP-SL	0.368932	-0.27781	0.069594	-0.52996	-0.21191
PRV-SL	-0.19591	**-0.65342**	0.491262	-0.0272	0.255816
PRA-SL	0.392578	**-0.64877**	-0.19824	0.094214	-0.11015
CPL-SL	-0.43658	**0.72541**	0.40134	-0.07598	-0.00195
CPD-SL	**0.857189**	-0.14573	0.210169	-0.03296	0.135031
CPD-CPL	**0.637875**	**-0.68363**	-0.2633	0.069884	0.023538
Proportion of total variation	0.258706	0.173578	0.144423	0.127219	0.072251

Figure 40. PCA based on percentage data, of the Zambezi River *Sargochromis*. This plot is between PC1 and PC2. N = 28. *Sargochromis carlottae* and *S. giardi* seprarate out together. *Sargochromis 'longsnout zambezi'* and '*Sargochromis- pharyngochromis'* have a close morphological similarity. *S. codringtonii* separates out alone. The variables with the highest factor loadings on PC1 and PC2 are set in bold.



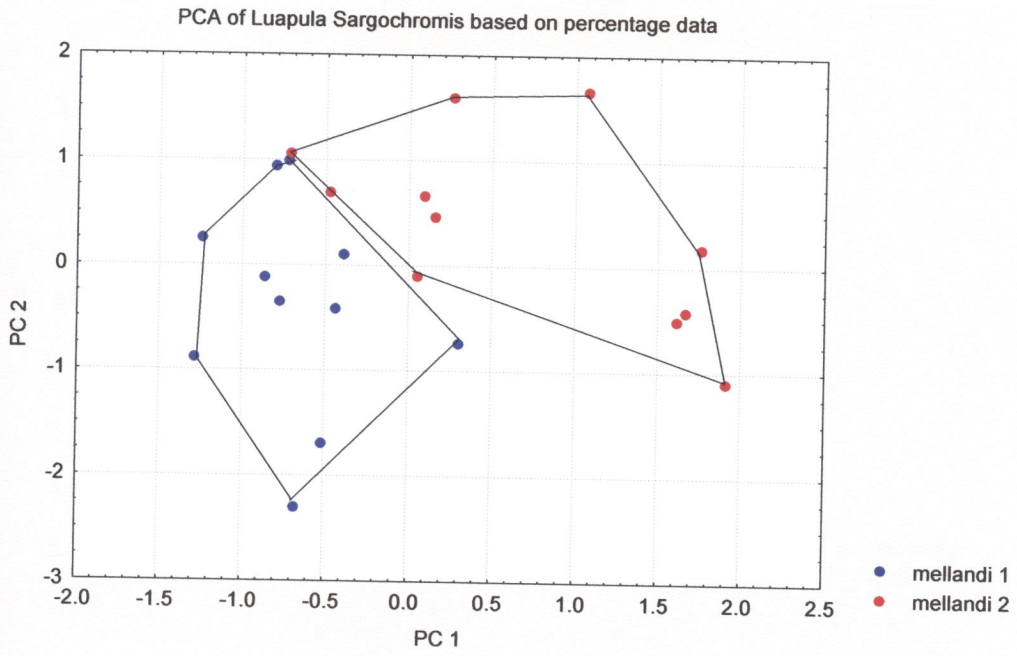
Variable	PC1	PC2	PC3	PC4	PC5
UJT	-0.60997	0.471714	0.023005	0.292536	0.405499
LJT	-0.62101	0.218867	0.11983	0.491437	0.362769
UJIR	-0.67022	0.379288	-0.05455	-0.17909	-0.07931
LJIR	-0.64396	0.525957	-0.24533	-0.1555	-0.00808
CERATO	-0.75974	0.033833	0.227151	-0.30089	-0.15101
EPI	0.244197	0.097867	-0.69419	-0.17644	-0.06126
DSPIN	0.03875	-0.0561	0.081786	-0.60244	0.641523
DSOFT	**0.635254**	0.130884	-0.06679	0.056138	-0.39084
ASOFT	0.219872	**0.600603**	0.36176	-0.17512	-0.20012
P	0.364933	**0.622264**	-0.35531	0.157248	0.008807
ULL	**0.499633**	-0.20466	0.007069	0.463391	0.223725
LLL	0.49238	0.095269	0.498922	-0.30268	0.229804
LOL	**0.891153**	-0.05239	0.19799	0.029407	0.047683
DLLA	0.012944	**0.675845**	-0.24483	-0.13246	-0.16853
ALLA	**0.712919**	0.37349	0.102409	-0.00099	0.011549
PV	0.34018	-0.07804	-0.73127	-0.10461	0.355322
CP	0.137187	**0.792677**	0.242193	0.175449	0.012203
CK	**0.67212**	**0.480667**	0.067623	-0.07144	0.271801
Proportion of total variation	0.29033	0.165849	0.102022	0.072116	0.070407

Figure 41. PCA based on meristic data, of the Zambezi River *Sargochromis*. This plot is between PC1 and PC2. N = 32. *Sargochromis giardi* separates from the rest on PC1. The variables with the highest factor loadings on PC1 and PC2 are set in bold. Tooth counts ar ignored.



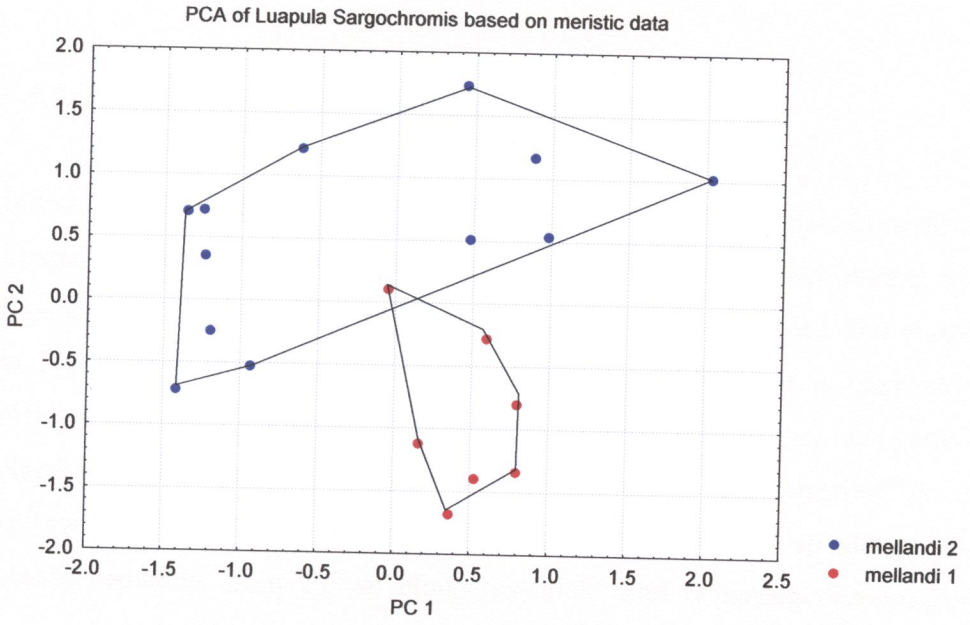
Variable	PC1	PC2	PC3	PC4	PC5
LOGLACRD	0.962269	0.043245	-0.06637	0.186854	-0.1338
LOGSNL	0.958962	-0.1026	-0.18327	0.047109	-0.07579
LOGLJL	0.966624	**-0.10389**	0.074934	-0.03729	-0.15232
LOGPPL	0.986544	0.025035	0.048681	-0.05938	0.028923
LOGCHD	0.974458	-0.04962	0.077238	-0.12742	0.011849
LOGED	0.923525	**-0.25919**	0.218244	0.157119	0.052341
LOGIOW	0.942198	**-0.15673**	-0.20583	0.061721	0.161196
LOGHW	0.988931	0.020687	-0.06992	-0.03992	-0.04848
LOGHL	0.993376	-0.03122	-0.03691	-0.03943	-0.02709
LOGSL	0.994755	0.055867	0.056235	-0.03944	0.029077
LOGBD	0.988857	0.041531	0.020519	-0.03377	-0.01956
LOGDFB	0.990983	0.048363	0.03554	-0.00255	0.02037
LOGAFB	0.972337	**0.118637**	-0.01382	-0.03142	0.02258
LOGPRD	0.993486	-0.02077	0.016758	-0.03185	0.039587
LOGPRP	0.992621	0.005876	0.023885	-0.04701	0.039661
LOGPRV	0.974357	-0.02736	-0.10339	-0.10608	0.018514
LOGPRA	0.993068	0.059818	0.052024	-0.04345	0.028404
LOGCPL	0.94319	**0.243219**	0.030942	0.173221	0.089026
LOGCPD	0.983672	0.072359	0.023987	0.036937	-0.07871
Proportion of total variation	0.950954	0.011046	0.008895	0.00744	0.005273

Figure 42. PCA based on log-transformed data, of the Luapula River *Sargochromis*. This plot is between PC1 and PC2. N = 22. The specimens are morphologically similar. The variables with the highest factor loadings on PC1 and PC2 are set in bold.



Variable	PC1	PC2	PC3	PC4	PC5
LACRD-HL	0.518199	-0.41456	0.128485	-0.26247	-0.45276
SNL-HL	0.10621	**-0.57634**	0.327254	-0.28106	0.190706
LJL-HL	0.48593	0.09235	-0.05808	0.405248	0.27079
PPL-HL	0.36121	**0.473506**	0.18566	-0.45362	0.456621
CHD-HL	**0.749476**	0.154205	-0.08414	0.263507	0.238625
ED-HL	**-0.68827**	0.1536	0.267025	0.005981	0.280412
IOW-HL	-0.18798	**-0.58544**	0.605498	0.223465	0.294034
IOW-HW	-0.44036	**-0.52446**	0.411312	0.438826	0.240301
HW-HL	0.587799	-0.04655	0.236912	-0.49269	0.058381
HL-SL	**-0.82846**	-0.29101	-0.22943	-0.00984	-0.10624
BD-SL	0.457855	-0.5016	-0.37571	0.033467	-0.04971
DFB-SL	0.614797	**-0.44081**	0.394866	-0.02825	-0.01125
AFB-SL	0.378853	-0.07264	0.216868	0.164428	-0.69622
PRD-SL	-0.63044	-0.42094	-0.27138	-0.28497	0.296008
PRP-SL	-0.6591	-0.32903	-0.41958	-0.26957	0.021537
PRV-SL	**-0.76133**	-0.33159	-0.30234	-0.13889	-0.09713
PRA-SL	**0.67371**	-0.11034	-0.15847	-0.4747	0.262271
CPL-SL	-0.48487	0.173327	0.523916	-0.37737	-0.27207
CPD-SL	**0.723031**	-0.38589	-0.10128	-0.11028	-0.06242
CPD-CPL	**0.776537**	-0.35231	-0.41776	0.192337	0.164286
Proportion of total variation	0.345017	0.132531	0.1027	0.084586	0.078809

Figure 43. PCA based on percentage data, of the Luapula River *Sargochromis*. This plot is between PC1 and PC2. N = 22. The specimens separate into two groups, with overlaps, on both PC1 and PC2. The variables with the highest factor loadings on PC1 and PC2 are set in bold.



Variable	PC1	PC2	PC3	PC4	PC5
UJT	0.065312	0.787097	0.106081	-0.22976	0.227364
LJT	-0.00518	0.776712	0.477051	-0.20383	0.147877
UJIR	0.477016	0.578034	-0.40641	0.090309	0.233133
LJIR	0.549483	0.394786	-0.53859	0.108728	0.290665
CERATO	0.74336	-0.23445	0.006604	0.303783	0.184362
EPI	-0.29828	-0.03813	0.346308	0.761186	0.328845
DSPIN	0.117064	0.071157	-0.70171	0.499936	-0.30883
DSOFT	0.412146	**-0.47088**	0.464088	0.112816	0.289029
ASOFT	**0.663279**	-0.12271	0.237351	-0.26246	0.102002
P	0.239191	-0.31475	0.722284	0.340345	-0.21808
ULL	0.411569	**0.42788**	0.373508	0.332786	-0.30302
LLL	**0.731658**	-0.33182	-0.27106	-0.30981	0.099693
LOL	**0.701771**	**0.414317**	-0.16802	0.312674	-0.27131
DLLA	0.408589	**-0.63906**	0.102537	-0.25259	-0.04599
ALLA	-0.40329	**-0.54548**	-0.49506	-0.05153	-0.13623
PV	**0.880514**	-0.10215	0.034751	-0.04632	-0.16866
CP	**0.743696**	-0.42356	-0.14363	0.070464	0.207819
CK	0.539055	0.347415	0.262674	-0.29478	-0.41268
Proportion of total variation	0.277878	0.199525	0.149306	0.094352	0.057188

Figure 44. PCA based on meristic data, of the Luapula River *Sargochromis*. This plot is between PC1 and PC2. N = 20. The specimens separate into two groups, with a minor overlap, on PC2. The variables with the highest factor loadings on PC1 and PC2 are set in bold.

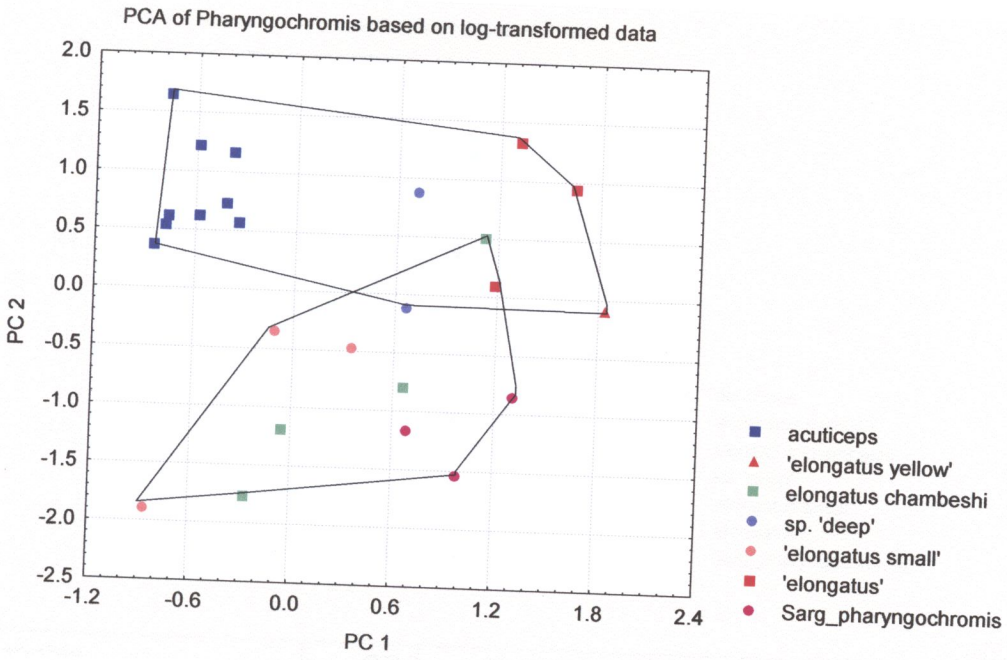
4.1.6. *Pharyngochromis* morphological results

Initial PCA results based on log-transformed data of 32 individuals from the Zambezi and Luapula systems show that there are two major clusters, which overlap as depicted in Figures 45a. The second PCA on log-transformed data idendtified four clusters as shown in Figure 45b. Based on the results of the second PCA, it is observed that *P.* 'sp elongatus-small', '*Pharyngochromis-Sargochromis*' and *P.* 'sp elongatus–chambeshi' clearly separate from the *Pharyngochromis acuticeps*/ *P.* 'sp elongatus'/ *P.* 'sp elongatus-yellow' cluster on PC2. It can be further observed that *P.* 'sp elongatus–chambeshi' clearly separates from *P.* 'sp elongatus-small' and '*Pharyngochromis-Sargochromis*' (which have a minor overlap with each other) on PC3. The variables with the highest loadings on PC2 were lachrymal depth (LacrD), eye diameter (ED), inter-orbital width (IOW), anal fin base (AFB) and dorsal fin base (DFB). On PC3 the variable with the highest loadings were lower jaw length (LJL), cheek depth (CHD), eye diameter (ED), interorbital width (IOW) and caudal peduncle length (CPL). The group that conforms to Skelton's (1993) description of *Pharyngochromis acuticeps* is the one collected from Lake Kariba and is identified as such in this analysis.

The results of the analysis based on percentage data of the same 32 individuals identified four clusters as depicted in Figure 46. *Pharyngochromis acuticeps* and *P.* 'sp elongatus-small' separate from the rest on PC1 while *P.* 'sp elongatus–chambeshi' separates from the rest on PC2. *P.* 'sp deep' with *P.* 'sp elongatus' separate from '*Pharyngochromis-Sargochromis*' and *P.* 'sp elongatus-yellow' on PC2. The variables with the highest loadings on PC1 were theratios; lachrymal depth (LacrD-HL), inter-orbital width (IOW-HW), anal fin base (AFB-SL) and caudal peduncle depth (CPD). On PC3, the variables with the highest loadings were the ratios; head length (HL-SL), pre-dorsal length (PRD-SL), pre-pelvic length (PRV-SL), pre-ventral length (PRV-SL) and caudal peduncle length (CPL-SL).

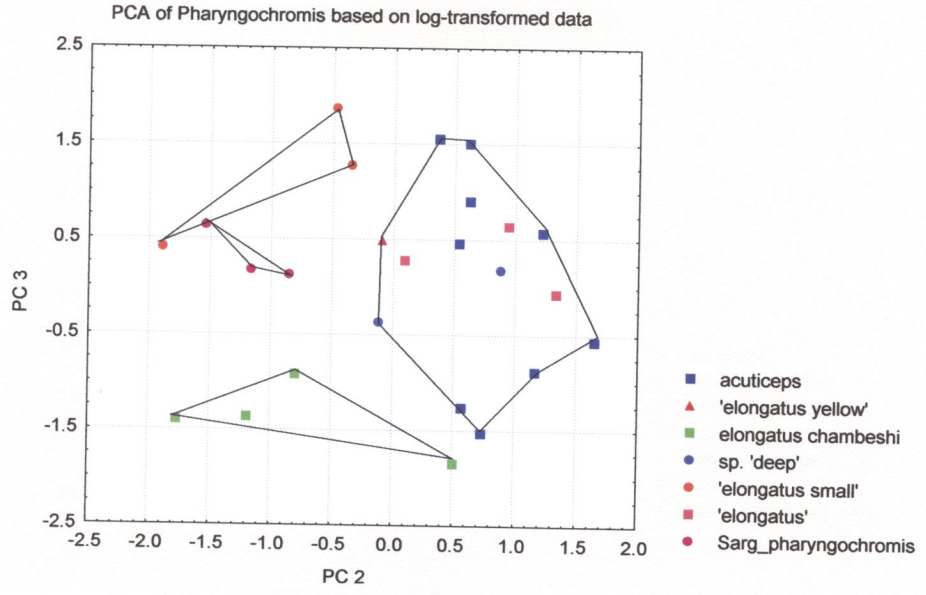
Figure 47 shows the PCA analysis of 33 individuals, based on meristic data in which three clusters are evident. *Pharyngochromis acuticeps* from Lake Kariba separates from

the rest, with a minor overlap, on PC1. The *Pharyngochromis* 'sp elongatus-chambeshi'/*P.* 'sp deep' cluster separates, with a minor overlap, from the *P.* 'sp elongatus'/*P.* 'sp elongatus-yellow'/P. 'sp elongatus-small'/ '*Pharyngochromis-Sargochromis*' cluster on PC2. The variables with the highest loadings on PC1 were ceretobranchial gill rakers (CERATO), dorsal and anal fin soft rays (DSOFT and ASOFT), pectoral fin rays (P) and scales between pelvic and ventral fins (PV). On PC2, the variables with the highest loadings were dorsal spiny rays (DSPIN) and lateral line scales (ULL, LLL and LOL).



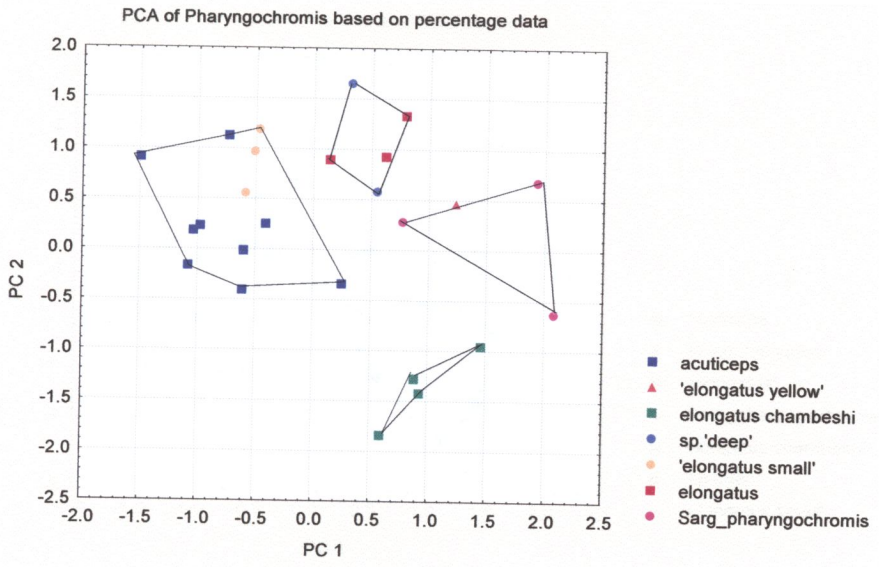
Variable	PC1	PC2	PC3	PC4	PC5
LOGLACRD	0.955685	**-0.17879**	-0.02151	0.07017	-0.19816
LOGSNL	0.98485	-0.03469	0.02752	0.028342	-0.09761
LOGLJL	0.970704	0.085083	-0.13001	0.089216	0.001072
LOGPPL	0.980305	-0.03409	0.034045	-0.06546	-0.00964
LOGCHD	0.958739	0.09043	0.087173	-0.16901	-0.06457
LOGED	0.935694	**0.263669**	-0.17338	0.102277	0.010968
LOGIOW	0.956661	**-0.19116**	-0.1316	0.047781	0.061837
LOGHW	0.99363	-0.00245	-0.04169	-0.01862	0.037409
LOGHL	0.994601	-0.02246	-0.01431	0.018927	-0.03834
LOGSL	0.995766	0.041447	0.058674	0.029948	0.027058
LOGBD	0.968187	-0.00164	0.010574	-0.15565	0.01567
LOGDFB	0.987917	**0.092167**	0.054032	-0.02179	0.029308
LOGAFB	0.951427	**-0.18158**	0.020845	0.035129	0.177091
LOGPRD	0.992818	-0.00952	-0.04978	-0.0242	-0.03155
LOGPRP	0.993361	-0.04588	-0.01703	0.011264	0.013246
LOGPRV	0.993242	-0.00673	-0.00083	-0.00695	-0.03321
LOGPRA	0.990329	0.081329	0.019903	-0.02645	0.02155
LOGCPL	0.936258	0.055997	0.269566	0.195155	0.025069
LOGCPD	0.98004	0.000871	-0.00031	-0.12744	0.055252
Proportion of total variation	0.95052	0.011131	0.008365	0.007505	0.005236

Figure 45a. PCA based on log-transformed data, of the *Pharyngochromis* from the Zambezi and Luapula systems. This plot is between PC1 and PC2. N = 32. Two clusters are evident, suggesting that *Pharyngochromis* is likely to comprise more than one species. The variables with the highest factor loadings on PC 2 are set in bold while PC1 variables are size related.



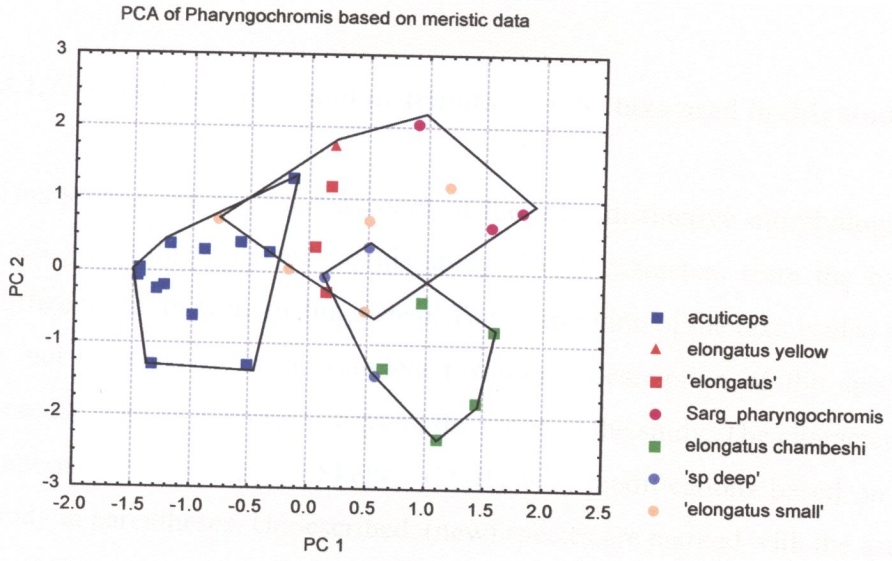
Variable	PC1	PC2	PC3	PC4	PC5
LOGLACRD	0.955685	**-0.17879**	-0.02151	0.07017	-0.19816
LOGSNL	0.98485	-0.03469	0.02752	0.028342	-0.09761
LOGLJL	0.970704	0.085083	**-0.13001**	0.089216	0.001072
LOGPPL	0.980305	-0.03409	0.034045	-0.06546	-0.00964
LOGCHD	0.958739	0.09043	**0.087173**	-0.16901	-0.06457
LOGED	0.935694	**0.263669**	**-0.17338**	0.102277	0.010968
LOGIOW	0.956661	**-0.19116**	**-0.1316**	0.047781	0.061837
LOGHW	0.99363	-0.00245	-0.04169	-0.01862	0.037409
LOGHL	0.994601	-0.02246	-0.01431	0.018927	-0.03834
LOGSL	0.995766	0.041447	0.058674	0.029948	0.027058
LOGBD	0.968187	-0.00164	0.010574	-0.15565	0.01567
LOGDFB	0.987917	**0.092167**	0.054032	-0.02179	0.029308
LOGAFB	0.951427	**-0.18158**	0.020845	0.035129	0.177091
LOGPRD	0.992818	-0.00952	-0.04978	-0.0242	-0.03155
LOGPRP	0.993361	-0.04588	-0.01703	0.011264	0.013246
LOGPRV	0.993242	-0.00673	-0.00083	-0.00695	-0.03321
LOGPRA	0.990329	0.081329	0.019903	-0.02645	0.02155
LOGCPL	0.936258	0.055997	**0.269566**	0.195155	0.025069
LOGCPD	0.98004	0.000871	-0.00031	-0.12744	0.055252
Proportion of total variation	0.95052	0.011131	0.008365	0.007505	0.005236

Figure 45b. PCA based on log-transformed data, of the *Pharyngochromis* from the Zambezi and Luapula systems. This plot is between PC2 and PC3. N = 32. *Pharyngochromis* 'sp elongatus' and *P. acuticeps* separate together just as *Pharyngochromis* 'sp elongatus-small' and '*Sargochromis-pharyngochromis*' do. *Pharyngochromis* 'sp elongatus-chambeshi' separates out alone. This result shows that *Pharyngochromis* is likely to comprise more than one species. The variables with the highest factor loadings on PC2 and PC3 are set in bold.



Variable	PC1	PC2	PC3	PC4	PC5
LACRD-HL	**0.715153**	0.144909	0.379145	-0.05209	0.077298
SNL-HL	0.604294	0.317464	0.128168	-0.26301	0.217248
LJL-HL	-0.03828	0.186864	-0.28207	0.750347	-0.08319
PPL-HL	0.334786	0.225657	-0.28796	-0.18303	0.697308
CHD-HL	0.193989	0.335167	-0.54485	-0.33778	0.160837
ED-HL	-0.71206	-0.15858	-0.42507	0.206522	-0.00202
IOW-HL	**0.846818**	0.1576	0.194161	0.356286	-0.11945
IOW-HW	**0.827491**	0.102933	0.351345	0.28294	-0.17707
HW-HL	0.18207	0.233412	-0.6864	0.40535	0.23704
HL-SL	0.289846	**-0.75742**	0.381868	-0.03587	-0.09737
BD-SL	0.613642	-0.17606	-0.22403	-0.32516	-0.20991
DFB-SL	0.470949	0.333561	-0.43865	-0.28965	-0.42316
AFB-SL	**0.712678**	0.161426	0.194883	0.225255	0.348706
PRD-SL	0.279979	**-0.77248**	-0.0897	0.053373	0.092646
PRP-SL	0.424772	**-0.65472**	0.163268	0.117923	0.292058
PRV-SL	0.075453	**-0.82596**	0.144841	-0.12345	0.093892
PRA-SL	-0.51316	-0.58288	-0.44193	-0.06651	0.102303
CPL-SL	-0.43615	**0.584837**	0.528989	-0.13876	0.064581
CPD-SL	**0.764211**	0.001176	-0.46393	-0.11792	-0.17179
CPD-CPL	0.704821	-0.33446	-0.55666	0.020824	-0.12578
Proportion of total variation	0.297947	0.183913	0.145741	0.075406	0.059544

Figure 46. PCA based on percentage data, of the *Pharyngochromis* from the Zambezi and Luapula systems. This plot is between PC1 and PC2. N = 32. Four clusters are evident. *P. acuticeps* and P. '*elongatus-small*' separate from the others on PC1. *Pharyngochromis* 'sp elongatus-chambeshi' cleary separates from the others on PC2. The variables with the highest factor loadings on PC1 and PC2 are set in bold.



Variable	PC1	PC2	PC3	PC4	PC5
UJT	-0.36814	0.080757	0.648626	0.398167	0.243315
LJT	-0.01248	-0.17751	0.492571	0.232712	0.562164
UJIR	-0.53029	-0.47895	0.240518	0.351998	0.034095
LJIR	-0.33514	-0.60754	0.183313	-0.1632	0.024191
CERATO	**0.662631**	-0.33859	-0.27604	-0.19549	0.18311
EPI	0.521856	-0.20882	0.184098	-0.29863	0.479525
DSPIN	-0.23767	**0.425311**	-0.2914	0.477143	0.296811
DSOFT	**0.759277**	-0.22777	0.137591	0.074549	-0.16027
ASOFT	**0.825611**	-0.03258	0.199347	0.0648	0.194455
P	**0.728386**	-0.25561	-0.1765	0.323896	-0.2359
ULL	0.522744	0.406521	-0.40822	0.117751	0.391306
LLL	0.15542	**0.681784**	0.582876	-0.08201	-0.09526
LOL	0.449928	**0.78529**	0.113008	-0.05726	-0.03638
DLLA	-0.04719	0.108875	-0.33288	0.680553	-0.3071
ALLA	-0.13848	0.079548	-0.44764	-0.18009	0.230442
PV	**-0.67132**	0.085377	-0.33783	-0.49172	0.102141
CP	0.380137	-0.5589	-0.17311	0.199558	-0.00669
CK	0.303887	0.014754	0.47478	-0.28833	-0.4282
Proportion of total variation	0.238587	0.148821	0.124416	0.095398	0.075137

Figure 47. PCA based on meristic data, of the *Pharyngochromis* from the Zambezi and Luapula systems. This plot is between PC1 and PC2. N = 33. Three clusters are evident. *P. acuticeps* from separates from the rest with a minor overlap on PC1. The two other cluster separate from each other with a minor overlap on PC2. This result shows that *Pharyngochromis* is likely to comprise more than one species. The variables with the highest factor loadings on PC1 and PC2 are set in bold. Tooth counts ar ignored.

4.1.7. Brief description and distribution of the taxa used in this study

This section gives a brief account of the most distinctive morphological characters for each group distinguished in this study. These characters form the basis on which the different groups were distinguished. The distribution of the taxa is also given. The section is not intended to be a complete taxonomic description of the species in this study because such an exercise is beyond the scope of the study. The characters and distribution patterns are according to Skelton (2001) with modifications based on results from this study in parentheses. Undescribed (new) species are marked with the asterisk (*).
A preliminary identification key is given under section 4.5.

4.1.7.1 Genus *Pseudocrenilabrus* Fowler

The genus *Pseudocrenilabrus* is widely distributed in Africa, from the northernmost edge of Egypt down to South Africa (Katongo *et al.*, *in press*). Within the broad African haplochromine group, the genus *Pseudocrenilabrus* has been placed in the H-lineage (Nishida, 1991and 1997; Salzburger et *al.*, 2002). Therefore, this genus forms part of Lake Tanganyika's primary lacustrine radiation, representing one of the oldest freshwater fish stocks. From the study area (Zambia) and the rest of South Eastern Africa, only one species, *Pseudocrenilabrus philander* has been reported (Skelton, 1991) and only one subspecies, *Pseudocrenilabrus philander philander* (Jackson, 1961; Skelton, 1993). Genetic analysis of this genus from the study area has revealed four major lineages, three occurring in the Congo drainage and one in the Zambezi drainage (Katongo *et al.*, *in press*).

The genus *Pseudocrenilabrus* is characterized by a small body size, an upper pharyngeal apophysis of the *Haplochromis*-type, a moderately protractile upper jaw, bicuspid and /or conical teeth in the outer row and tricuspid teeth in the inner rows of both jaws, ctenoid scales, an interrupted single pored lateral line and a rounded caudal fin as an autapomorphy. The males lack egg spots in the anal fin, displaying a distinct red or orange spot on the distal tip of the anal fin. They are maternal brooders with up to 120 fry per clutch. The genus comprises three nominal species even if the actual taxonomic

situation is more complex (Greenwood, 1989). The observed species diversity of the genus *Pseudocrenilabrus* is comparatively lower than that for the other riverine cichlid genera of Zambia (Skelton, 1991).

4.1.7.1.1. *Pseudocrenilabrus philander* (Weber, 1897)

Description: D XIII (VX), 9-11; A III, 7-9(10). Lateral line scales (25) 27-30, chest scales not sharply differentiated from body scales. Body stout, caudal fin rounded. Mouth small, horizontal. Male colours differ with locality; body usually a mesh of iridescent light blue and yellow, with an oblique bar through the eye and iridescent blue lower jaw; dorsal fin has a red tip, black submarginal band and iridescent blue blocks. Pelvics black, caudal and anal fins have iridescent blue and red blocks, anal fin with a red tip. Attains 130 mm TL.

Distribution: Side arms and quiet waters of the Lukanga Swamps, Kabala (upper Zambezi River), Lake Kariba, Mukula Stream (near Lake Bangweulu) and Chambeshi River.

4.1.7.1.2. *Pseudocrenilabrus philander* 'blue' *P. philander* 'yellow' *

Description: D XIII-VX, 8-10; A III, 7-8. Lateral line scales 26-29.

Locality: Lunzua River bridge.

4.1.7.1.3. *Pseudocrenilabrus philander* 'sp lufubu' *

Description: D XIII-VX, 8-10; A III, 7-8. Lateral line scales 24-30.

Distribution: Lufubu River system.

4.1.7.1.4. *Pseudocrenilabrus* ' sp haplochromis-orange' *

Diagnosis: Pale and fragile-looking *Pseudocrenilabrus* from Lake Mweru.

Description: D XIII-VX, 9-10; A III, 6-9. Lateral line scales 28-32. It is present in two colours, orange and blue.

Locality: Mwatishi River-Lake Mweru confluence.

4.1.7.2. Serranochromine cichlid group

This informally recognised, fluviate southern African cichlid group comprises the genera *Pharyngochromis* Greenwood, *Sargochromis* Regan, *Serranochromis* Regan and *Chetia* Trewavas. It is a group of non-tilapiine and non-pelmatochromine cichlids (Greenwood, 1993).

Morphologically, the serranochromines have two distinguishing features. They have maculae (anal markings) on the anal fin, usually in both sexes. These maculae are different from true ocelli which are only found in males of haplochromine species such as those of Lake Victoria. The second feature is based on squamation pattern. All scales above the lateral-line series are cycloid, as are the majority of scales below that level. Some weakly ctenoid scales may occur anteriorly on the flanks, especially in small specimens, the ctenii being confined to a narrow arc situated near the centre of the scales's free margin (Greenwood, 1993).

The monophyletic origin of the serranochromines is still to be established. The inter relationships of the serranochromine genera are based on shared derived characters which are not necessarily truly phylogenetic. The serranochromine group does not constitute a formal cichlid Tribe of their own as given by Poll (1986) to several cichlid assemblages in Lake Tanganyika. They belong to the tribe Haplochromini defined by Eccles and Trewavas (1989).

4.1.7.3. The Genus *Serranochromis* Regan, 1920

These are largemouth breams, a distinct group of predatory cichlids found in rivers of tropical southern and central Africa. They are easily recognized by their large mouths with conical teeth, the high number of scale rows on the cheeks, the relatively high number of soft rays in the dorsal fin and often by numerous bright egg-spots (ocelli) on the anal fin. They are found in the Zambezi River drainage, the Kafue River drainage and the Chambeshi/ Luapula River drainage including Lakes Mweru and Bangweulu.

4.1.7.3.1. *Serranochromis longimanus* (Boulenger, 1911)

Diagnosis: Distinguished from other species of *Serranochromis* by its exceptionally long pectoral fin.

Description: D XIV-XV (XVI), (12) 13-14; A III, (8)9-10; lateral line scales (33-37) 35; pectoral fin exceptionally long, 33-40 % SL; caudal fin rounded; mouth large (and very protractile); head acutely pointed, with large black blotches; 5-6 broad black bars on body; fins sooty with black blotches; teeth coarse.

Distribution: This species is distributed within the upper Zambezi floodplains.

4.1.7.3.2. *Serranochromis robustus* (Günter, 1864)

Diagnosis: The colour pattern has longitudinal elements predominating. The body is heavy-set and robust.

Description: D XV-XVI(XVII), 13-15(16); A III, 10-11(12). Lateral line scales 36-39(40). Heavy-set robust body, especially of large specimens, is characteristic; pectoral fins are relatively short, 19-23% SL, and rounded. Mouth large, with large well-spaced conical teeth. Body colours olive to bright green with yellow-orange margins.

Distribution: Upper Zambezi, Kafue and Luapula River systems.

4.1.7.3.3. *Serranochromis thumbergi* (Castelnau, 1861)

Diagnosis: Superficially similar to *S. robustus*. More slender than *S.robustus* with emarginated tail fin with less yellow or gold in body colouration. The colour pattern emphasizes the longitudinal elements.

Description: D (XV) XVII-XVIII, (12) 13-16; A III, (8)9-12; lateral line scales (37-44)39-41; a rust red spot is prominent at the base of each flank scale; depth of body nearly equal to length of head, 2.5 – 3 times in total length.

Distribution: Upper Zambezi, Kafue and Luapula River systems.

4.1.7.3.4. *Serranochromis angusticeps* (Boulenger, 1907)

Diagnosis: Head and body greatly compressed. Speckled mouth, cheek and gill cover is a diagnostic colour pattern (red spots and vermiculations) not present in the other members of this genus. Very deep body.

Description: D XIII-XVI, (13)14-17; A III, 11-13(15). Lateral line scales (34)36-39. Head and body greatly compressed, mouth very large, inclined 40-55°, with relatively fine conical teeth. Head and body of both sexes with distinctive red spots and vermiculations.

Distribution: Upper Zambezi, Kafue and Luapula River systems.

4.1.7.3.5. *Serranochromis altus* Winemiller & Kelso-Winemiller, 1990

Diagnosis: A species of *Serranochromis* that greatly resembles *S. angusticeps* in having a very deep body but without red spots and vermiculations. Head profile has a characteristic deep kink

Description: D XV-XVI, (13)15-18; A III, 11-13(15). Lateral line scales (34)37-39. Teeth close set in 4-6 rows. Mouth large and extremely protrusible.

Distribution: Upper Zambezi and Kafue River systems in sympatry with *S. angusticeps.*

4.1.7.3.6. *Serranochromis stappersi* Trewavas, 1964

Diagnosis: A *Serranochromis* that resembles *Serranochromis macrocephalus* but has very small teeth, almost hidden in the lips.

Description: D XIV-XV, 13-15; A III, 8-11. Lateral line scales 34-37. Mouth set at about 40-50° to the horizontal of the maxillary extending to below the eye and the pre-maxillary pedicel to between the orbits. Premaxillary pedicel 35.5-42% of head length and pectoral fin 31-34 % of standard length. Interorbital width 19-22.5 % of head length.

Distribution: Lake Mweru and Luapula River system.

4.1.7.3.7. *Serranochromis macrocephalus* (Boulenger, 1899)

Diagnosis: Straight head profile, large mouth with large well-spaced conical teeth and stocky build are characteristic.

Description: D XIV-XVI, (12)13-15; A III, (7)9-11. Lateral line scales (31)34-37. Straight head profile, large mouth with large well-spaced conical teeth and stocky build. Pectoral fins relatively long, 23-29% SL. Colour variable, vertical body bars usually evident.

Distribution: Upper Zambezi and Kafue systems, Lake Kariba. Also some tributaries of the Luapula system.

4.1.7.3.8. *Serranochromis* 'sp elongate'*

Diagnosis: Resembles *Serranochromis stappersi* but has a more elongate body profile.

Description: D XIV-XVI, 12-15; A III, 9-11. Lateral line scales 34-36. Upper row teeth 26-52; lower jaw teeth 22-32. Elongate body.

Locality: Mifimbo area, Luapula River-Lake Mweru estuary.

4.1.7.4. The Genus *Sargochromis* Regan, 1920

Fairly large, stout-bodied serranochromines, with moderate sized mouths, feeding mainly on snails, freshwater mussels and crustaceans. They are mostly characterized by robust pharyngeal bones with rounded, molar-like teeth. The females mouth brood the eggs and larvae, and the species are of importance to floodplain fisheries. As in other cichlid groups, the differences between the species are sometimes obscure externally and therefore not easy to identify positively. This is particularly true of the juveniles of these sargos. The last 2 or 3 pored scales in the upper lateral line series are separated from the dorsal fin base by not less than 2 scales of approximately equal size.

4.1.7.4.1. *Sargochromis carlottae* (Boulenger, 1905)

Diagnosis : IN adults the head has a distinctly concave profile.

Description : D XIV-XVI , (10)11-13; A III, (8)9-11.Lateral line scales 28-32(35). Body depth greater than head length, caudal fin rounded.

Distribution : Upper Zambezi and Kafue systems. Also in Lake Kariba where it has been introduced to impoundments in middle Zambezi catchment.

4.1.7.4.2. *Sargochromis codringtonii* (Boulenger, 1908)

Diagnosis : Head with straight profile, deep cheek and snout, mouth slightly inclined.

Description : DXIII-XVI, 11-16; AIII, (8)9-10. Lateral line scales 28-33(35). Body depth greater than head length.

Distribution : Upper Zambezi, Lake Kariba and Kafue systems.

4.1.7.4.3. *Sargochromis giardi* (Pellegrin, 1905)

Diagnosis : Robust big body and large head with a typical steep convex profile are characteristic.

Description : D XIV-XVI , 12-13(15); A III, 9-11. Lateral line scales 29-34(36). Mouth inclined, teeth stout, conical, in 2 rows.

Distribution : Upper Zambezi and Kafue systems and Lake Kariba.

4.1.7.4.4. *Sargochromis coulteri* (Bell-Cross, 1975)

Diagnosis : Deep bodied, caudal sub-truncate.

Description : DXIV-XVI, 11-13 ; A III, 8-9. Lateral line scales 30-31(36). Head large, mouth inclined, two rows of conical teeth in jaws.

Distribution : Cunene system (Chunga Lagoon, Kafue River).

4.1.7.4.5. *Sargochromis mellandi* Regan, 1922

Diagnosis : A *Sargochromis* species confined to the Luapula River system. Distinguished from the tilapias in that the body is slightly more elongate and the colour paler often with a hint of mauve.

Description : D XIV-XVI, 11-13; A III, 8-9. Lateral line scales 33-36. Dorsal spines equal in length from sixth or seventh. Depth of body nearly equal to length of head, 2.5 – 3 times in total length. Head a little more than twice as long as broad.

Distribution : Lake Bangweulu , Lake Mweru, Chambeshi River, and Lukulu River.

4.1.7.4.6. *Sargochromis* 'sp longsnout-kafue' *

Diagnosis : A *Sargochromis* characterised by a small adult size, a straight body profile and an elongate snout, morphologically similar to *S. mellandi*.

Description : D XIV-XV, 12-14; A III, 8-9. Lateral line scales 32-35.

Locality : Lukanga Swamps and Chanyanya Lagoon, Kafue River system

4.1.7.4.7. *Sargochromis* ' sp zambia' *

Description : D XV, 12-14; A III, 8-9. Lateral line scales 31-35.

Locality: Below Ithezhi-tezhi dam, Kafue River.

4.1.7.4.8. *Sargochromis* ' sp double striped1'*

Diagnosis : A *Sargochromis* with 2 longitudinal stripes along each side of the body.

Description : D XIV-XV, 10-12; A III, 8-9. Lateral line scales 32-35.

Locality: Kafue Flats, Kafue River.

4.1.7.4.9. *Sargochromis* ' sp codringtonii-small'*

Diagnosis : Small size with a straight body profile.

Description: D XV, 11; A III, 9. Lateral line scales 34-36.

Locality: Kafue Flats, Kafue River.

4.1.7.4.10. *Sargochromis* ' sp codringtonii-big'*

Description: D XV, 11; A III, 8- 9. Lateral line scales 33-34.

Locality: Kafue Flats and Lukanga Swamps, Kafue River.

4.1.7.4.11 '*Pharyngochromis-Sargochromis*' species *

Diagnosis : Has some characters of both *Sargochromis* and *Pharyngochromis*.

Description: D XV, 11; A III, 8-9. Lateral line scales 34-35.

Locality: Ngweze fishing camp, Upper Zambezi River.

4.1.7.5. The Genus *Pharyngochromis* Greenwood, 1979

This genus has a characteristic stout pharyngeal bone with coarse molariform or submolariform teeth. Males have numerous bright egg-spots on the anal fin. Last 5 to 7 pored scales in the upper lateral line series separated from the dorsal fin base by one large and one small scale.

4.1.7.5.1. *Pharyngochromis acuticeps* (Steindachner, 1866)

Description: D (XIV-XV) XIV-XVI, (8-10) 10-13; A III, (6-8) 7-10. Lateral line scales (31-33) 31-36, cheek scales (3-4) 4-5 rows. Body robust, caudal fin truncate. Head large, mouth horizontal or slightly oblique, conical and bicuspid teeth in 3 rows. Gill rakers 7-12, stout and T-shaped. Brown with black bars and iridescent green infusions, body scales with deep red centers; dorsal fin with sooty band, orange-red tips and sooty orange

spots especially on soft-rayed section, caudal fin with dark brown spots and the anal fin with orange-red margin and a variable number of clear orange egg-spots. Chest of breeding males dark sooty grey. Attains 220 mm TL, usually less than 100 mmTL.
Distribution: Zambezi system.

4.1.7.5.1. *Pharyngochromis* 'sp elongatus' *

Description: D XIV-XVI, 10-11; A III, 7-8. Lateral line scales 33.
Locality: Kabala and Ngweze fishing camps, Upper Zambezi River.

4.1.7.5.2. *Pharyngochromis* 'sp elongate small-chambeshi' *

Description: D XIV-XV, 10-12; A III, 8-9. Lateral line scales 30-33.
Locality: Chambeshi River Bridge.

4.1.7.5.3. *Pharyngochromis* 'sp elongatus small' *

Description: D XIV-XV, 10-12; A III, 7-8. Lateral line scales 32-34.
Locality: Kabala and Ngweze fishing camps, Upper Zambezi River.

4.2. Molecular Biological Analyses

The common strategy used in the analysis of both D-loop and ND2 data of the Zambian *Pseudocrenilabrus* species complex and the serranochromine cichlids involved Neighbour Joining with bootstrapping, Maximum Parsimony with bootstraping (optional), Maximum Likelihood with Puzzle (optional), 4-Cluster Likelihood Mapping analysis, Average Distances and Linearized Tree analysis.

4.2.1 *Pseudocrenilabrus* species complex - mtDNA phylogeny

The sample size used in this analysis is not as representative as that used for the serranochromines because individuals of this taxon were present only in a few of the sampling sites visited.

4.2.1.1. *Pseudocrenilabrus* D- Loop phylogeny

4.2.1.1.1 *Pseudocrenilabrus* D-Loop Neighbour Joining (NJ) tree

The NJ tree was constructed based on the TrN + I+ G model (Tamura and Nei, 1993) with a gamma distribution shape parameter (G) of 0.8794 and the proportion of invariable sites (I) being 0.3431.The base frequencies were; A (0.2907), C (0.3637), G (0.0894) and T (0.2561). The NJ tree with uncorrected p-distances shows identical topology and supports the same four clades (Figure 48a) as the NJ-tree based upon TrN + I+ G distances (Figure 48b). The individuals studied were the Lunzua population, Lake Tanganyika drainage (11 specimens); Lukanga Swamps population, Kafue River (three specimens); Mambova population, Upper Zambezi (two specimens); Siavonga population, Lake Kariba (four specimens); Mwatishi population, Lake Mweru (four specimens); Mukula population, Lake Bangweulu (five specimens) and the Chambeshi population, Luapula drainage (five specimens). *Pseudocrenilabrus nicholsi* (two specimens), and *Pseudocrenilabrus multicolour* (two specimens), were used as out-group taxa. The Bangweulu/Chambeshi population occupies the most ancestral branch in the NJ tree.

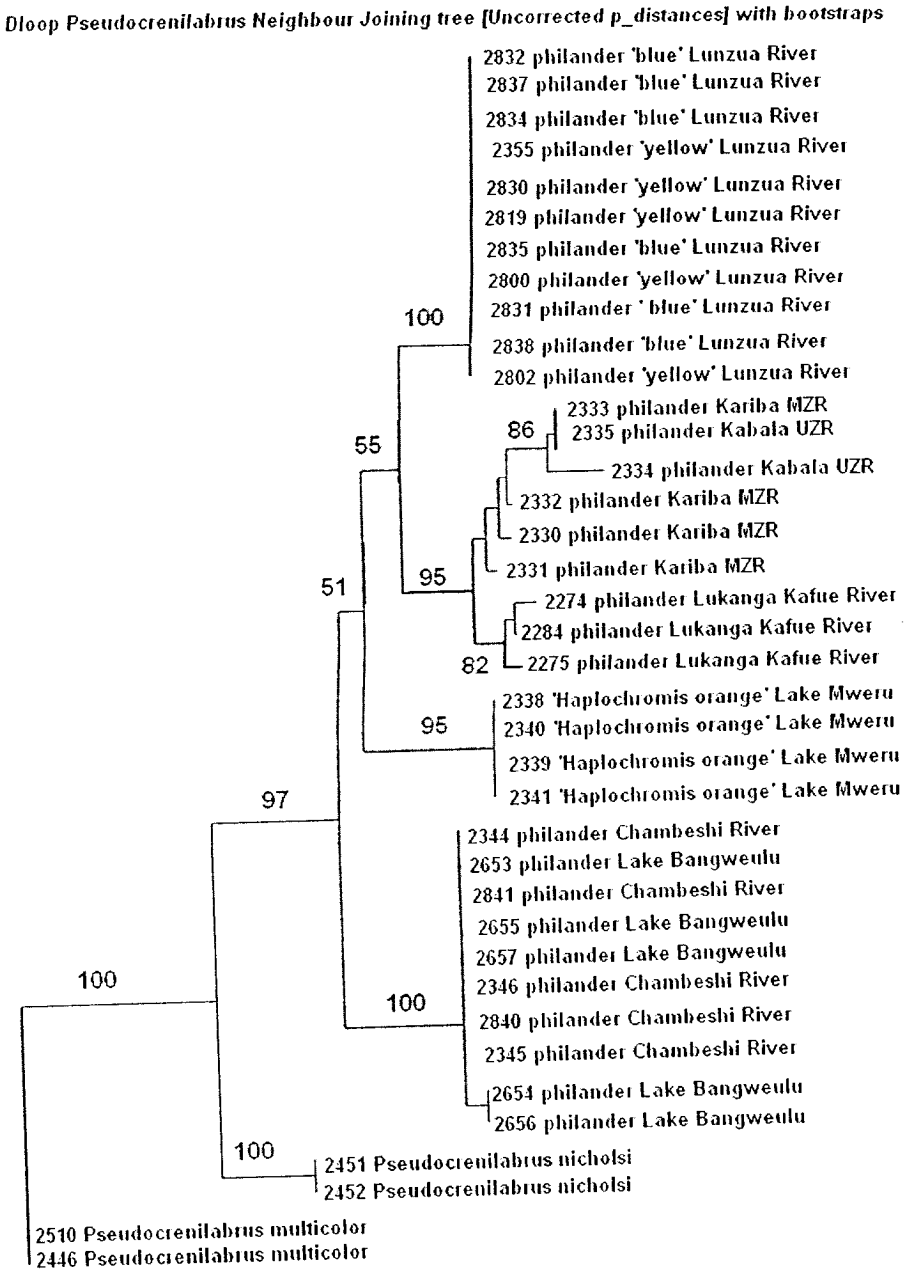


Figure 48a. Neighbour joining tree (uncorrected p-distances) of 38 taxa of the Zambian *Pseudocrenilabrus* species complex from seven localities, plus the two outgroup taxa *Pseudocrenilabrus multicolor* and *P. nicholsi*, based upon 358 bp of the most variable part of the control region (Figure 49). Only bootstraps larger than 50 are shown.

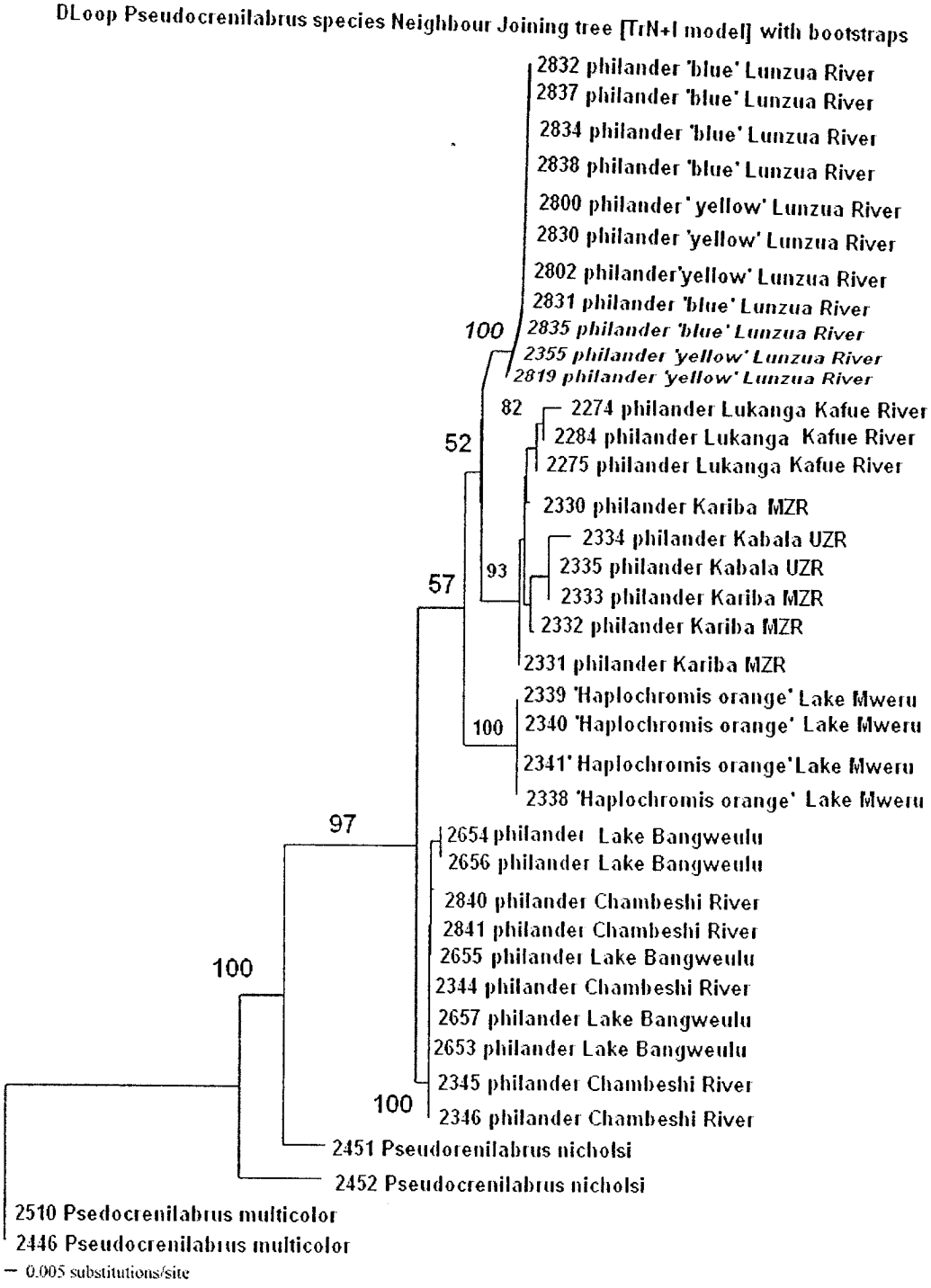


Figure 48b. Neighbour joining tree [substitution model TrN+I (Tamura and Nei, 1993)] of 38 taxa of the Zambian *Pseudocrenilabrus* species complex from seven localities, plus the two outgroup taxa *Pseudocrenilabrus multicolor* and *P. nicholsi*, based upon 358 bp of the most variable part of the control region (Figure 49). Only bootstraps larger than 50 are shown.

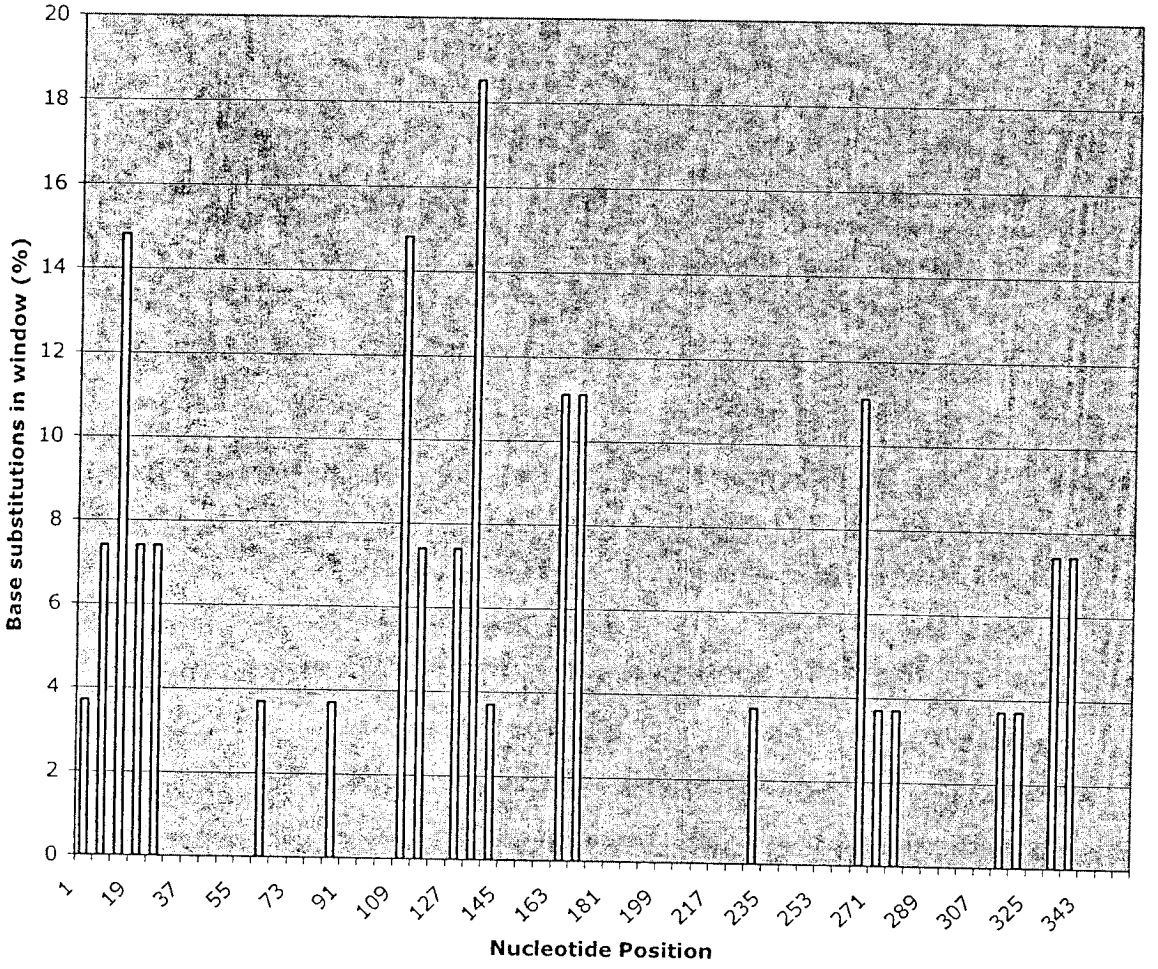


Figure 49. Sliding window analysis based on a 358 bp segment of the most variable part of the mitochondrial control region. Three regions were defined as showing low variability (Low Var) if the overall base substitution was less than ten percent (10%), high variability (High Var) if the overall base substitution was between ten and twenty percent (10-20%) and hyper variability (Hyper Var) if the overall base substitution was more than twenty percent (20%).

4.2.1.1.2 *Pseudocrenilabrus* D-Loop Unweigheted Maximim Parsimony (MP) tree

The strict consensus tree of two unweighted MP trees has 246 steps [consistency index (CI) excluding uninformative sites, 0.82; retention index (RI), 0.96; rescaled consistency index (RC), 0.79; (See Figure 50)]. All transitions (TS), TV transversions and insertions/deletions (indels) are given the same weight. The same individuals used in the calculation of the NJ trees, were used for the MP tree. The MP tree highly supports four clades but places the Lake Mweru/Mwatishi River population at the most ancestral split. Thus, there is some conflict concerning the most ancestral split in relation to the NJ tree (Figures 48a and 48b).

4.2.1.1.3 *Pseudocrenilabrus* D-Loop Maximum Likelihood (ML) tree

The ML tree was calculated using the HKY + G model (Hasegawa *et al.*, 1985) with a gamma distribution shape parameter (G) of 0.0145. Base frequencies were; A (0.3488), C (0.1889), G (0.1370) and T (0.3253). The ML tree again highly supports four clades, resolving the Lake Mweru population as the most ancestral split like in the MP analysis (Figure 51). The same individuals used in the calculation of the NJ and MP trees, were used for the ML tree.

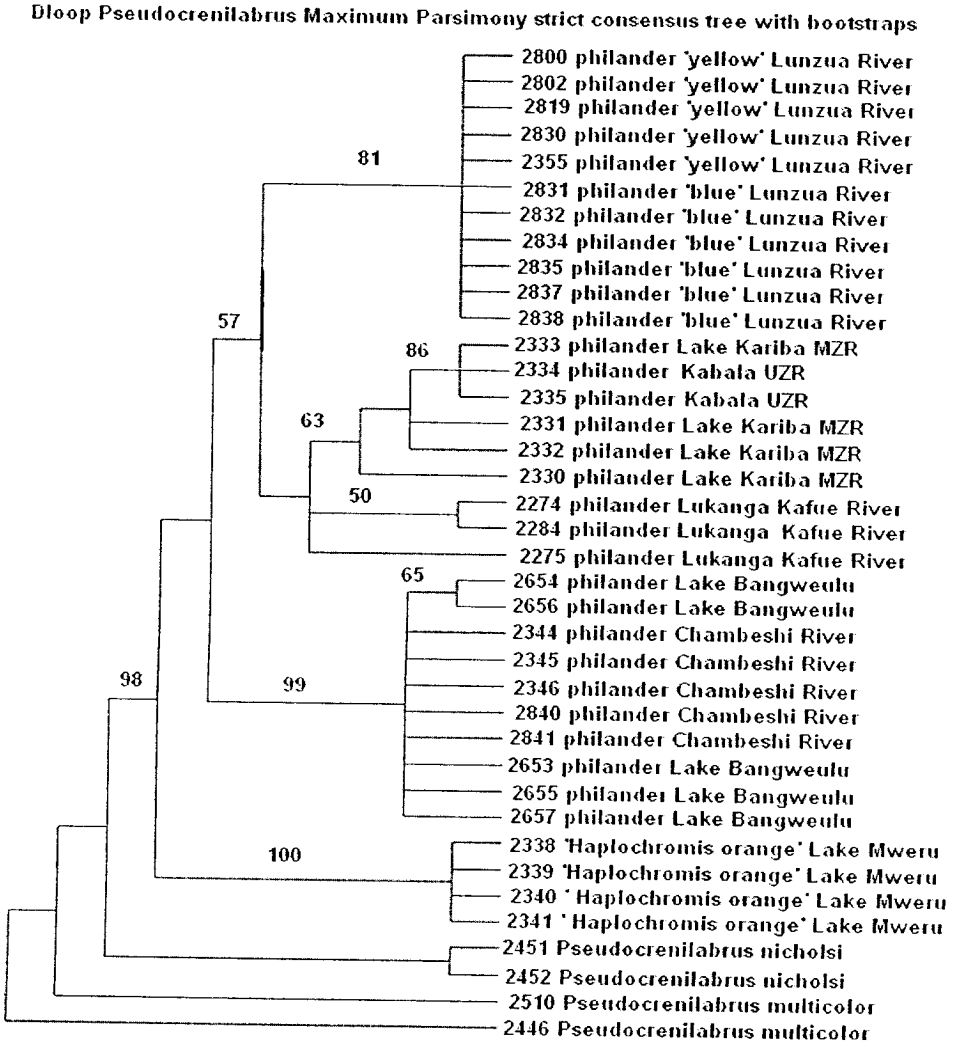


Figure 50. Strict consensus tree of two MP trees [246 steps; CI excluding uninformative sites, 0.82; retention index (RI), 0.96; rescaled consistency index (RC), 0.79], of 38 taxa of the Zambian *Pseudocrenilabrus* species complex from seven localities, plus the two outgroup taxa *Pseudocrenilabrus multicolor* and *P. nicholsi*, based upon 358 bp of the most variable part of the control region. Only bootstraps equal to or larger than 50 are shown.

Pseudocrenilabrus Maximum Likelihood tree [Substitution model HKY+G (Hasegawa et al.,1985)]

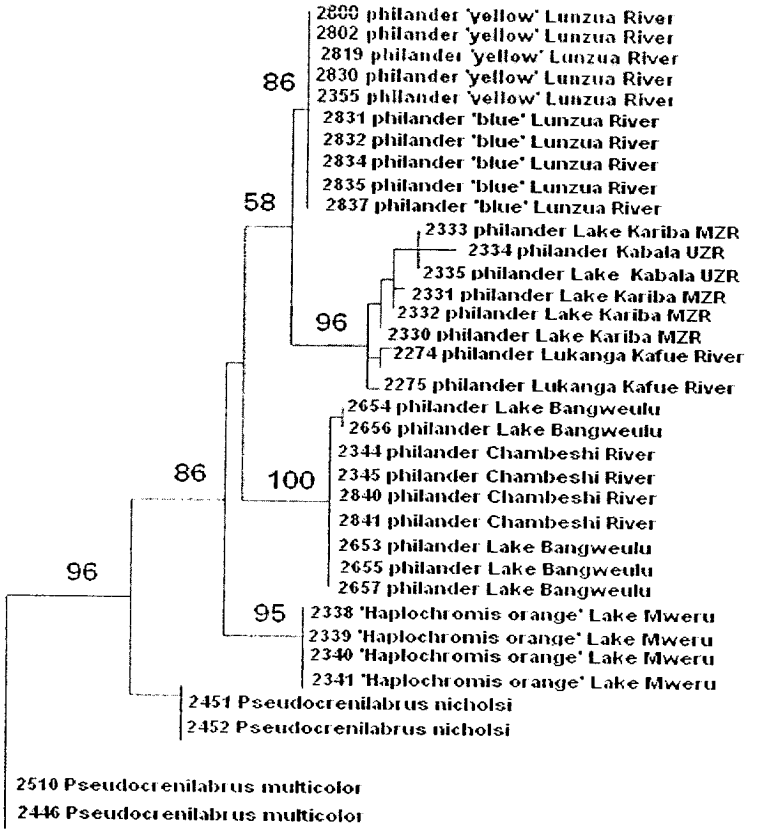


Figure 51. *Pseudocrenilabrus* Maximum Likelihood tree [Substitution model HKY+G (Hasegawa *et al.*, 1985)] of 38 taxa of the Zambian *Pseudocrenilabrus* species complex from seven localities, plus two outgroup taxa *Pseudocrenilabrus multicolor* and *P. nicholsi* based on 358bp of the most variable part of the mitochondrial control region. Only puzzle values equal to or greater than 50 are shown.

4.2.1.1.4 *Pseudocrenilabrus* Four Cluster Likelihood Mapping Analysis

This analysis gives very clear biogeographic pattern and phylogeographic affinities. The order of colonization is unclear. There is a strong support of bifurcation in one direction. (a,c)– (bd): Chambeshi/Bangweulu//Kafue/Upper Zambezi + Mweru /Lunzua. The clades are defined as follows: a = Chambeshi/Bangweulu; b = Mweru; c = Kafue/ Zambezi; d = Lunzua.

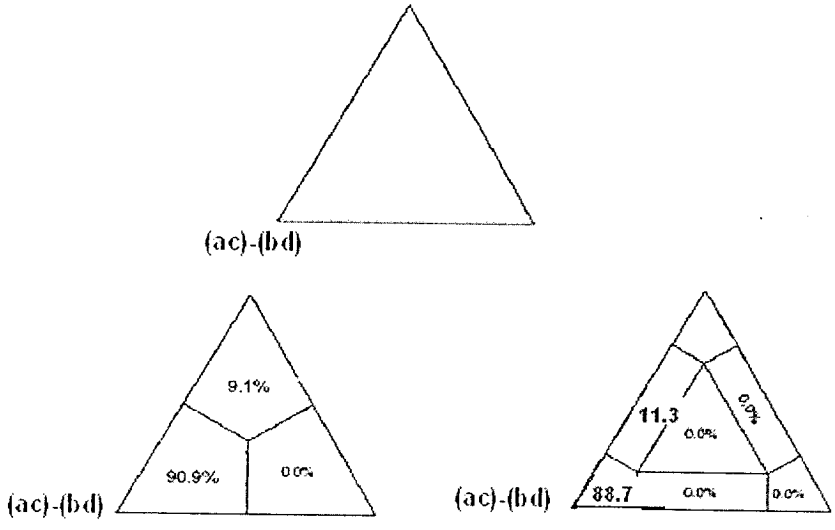


Figure 52. Four-Cluster Likelihood Mapping analysis of Zambian *Pseudocrenilabrus* species complex based on the most variable part of the mitochondrial control region. The clusters are Bangweulu and Chambeshi populations (a), Lake Mweru population (b), Zambezi and Kafue populations (c). The Lunzua population forms cluster d. This analysis strongly supports the (ac) – (bd) topology.

4.2.1.1.5 *Pseudocrenilabrus* Linearized tree

The linearized tree was constructed based on the HKY model [Hasegawa *et al.*, 1985] with each population represented by one individual. The linearized tree points to the contemporaneous evolution within one diversification event – one fast spread. A second and very recent diversifiication concerns the Kafue and Zambezi River populations (Figure 53).

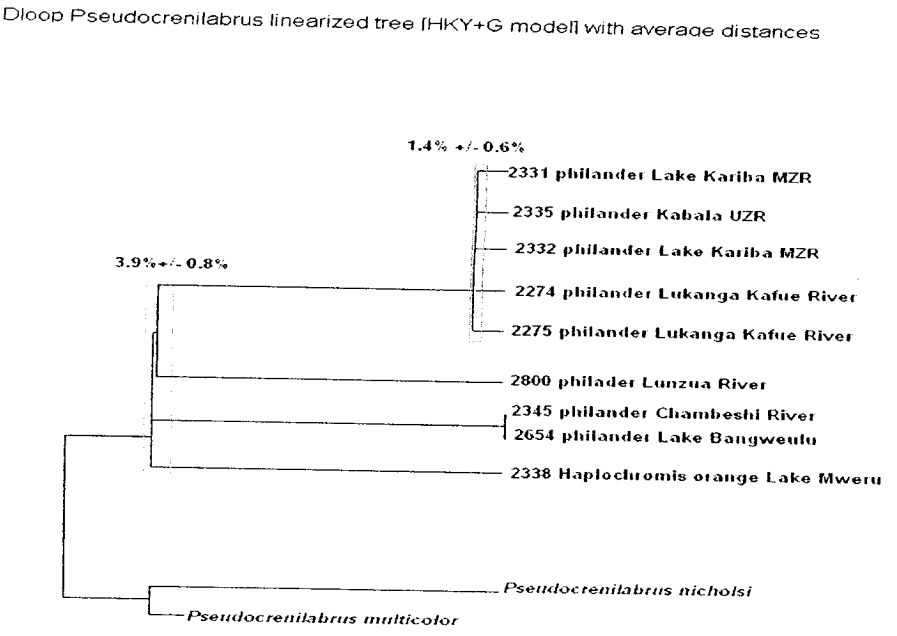


Figure 53. *Pseudocrenilabrus* Linearized tree based on a 358bp segment of the most variable part of the mitochondrial control region. The linearized tree was compiled with the computer program LINTRE (Takezaki *et al.*, 1995) after performing a branch length test (Takezaki *et al.*, 1995). Grey rectangles refer to the major diversification events. The observed mean square divergences using the substitution model 2KP (Kimura, 1980) are depicted as percentages for the corresponding diversification events.

4.2.1.2 *Pseudocrenilabrus* ND2 phylogeny

4.2.1.2.1 *Pseudocrenilabrus* ND2 Neighbour Joining (NJ) tree

The NJ tree was constructed based on the HKY + G model (Hasegawa *et. al.*, 1985) with a gamma distribution shape parameter (G) of 0.2705. The base frequencies were; A (0.2709), C (0.3406), G (0.1069 and T (0.2744). The tree is depicted in Figure 54.

The individuals studied were the Lunzua population, Lake Tanganyika drainage (four specimens); Lukanga Swamps population, Kafue River (two specimens); Mambova population, Upper Zambezi (two specimens); Siavonga population, Lake Kariba (two specimens); Mwatishi population, Lake Mweru (two specimens); Mukula population, Lake Bangweulu (two specimens) and the Chambeshi population, Luapula drainage (four specimens). *Pseudocrenilabrus nicholsi* (two specimens), and *Pseudocrenilabrus multicolour* (three specimens), were used as out-group taxa. The Lukanga Swamps population (from the Zambezi/Kafue drainage) forms the most ancestral split.

4.2.1.2.2 *Pseudocrenilabrus* ND2 Maximim Parsimony tree

The strict consensus tree of two Maximum Parsimony (MP) trees is depicted in Figure 55. The same individuals used in the calculation of the NJ trees, were used for the MP tree. The MP tree highly supports four clades with the Lukanga Swamps population forming the most ancestral split.

4.2.1.2.3 *Pseudocrenilabrus* ND2 Maximum Likelihood (ML)tree

The ML tree was calculated using the HKY +G model (Hasegawa *et al.*, 1985) with a gamma distribution shape parameter (G) of 0.2705. Base frequencies were: A (0.2781), C (0.3406), G (0.1069) and T (0.2744). See figure 56. The same individuals used in the calculation of the NJ trees, were used for the MP tree. The LH tree highly supports four clades with the Lukanga Swamps population forming the most ancestral split.

ND2 Pseudocrenilabrus Neighbour joining tree [HKY+G Model] with bootstraps

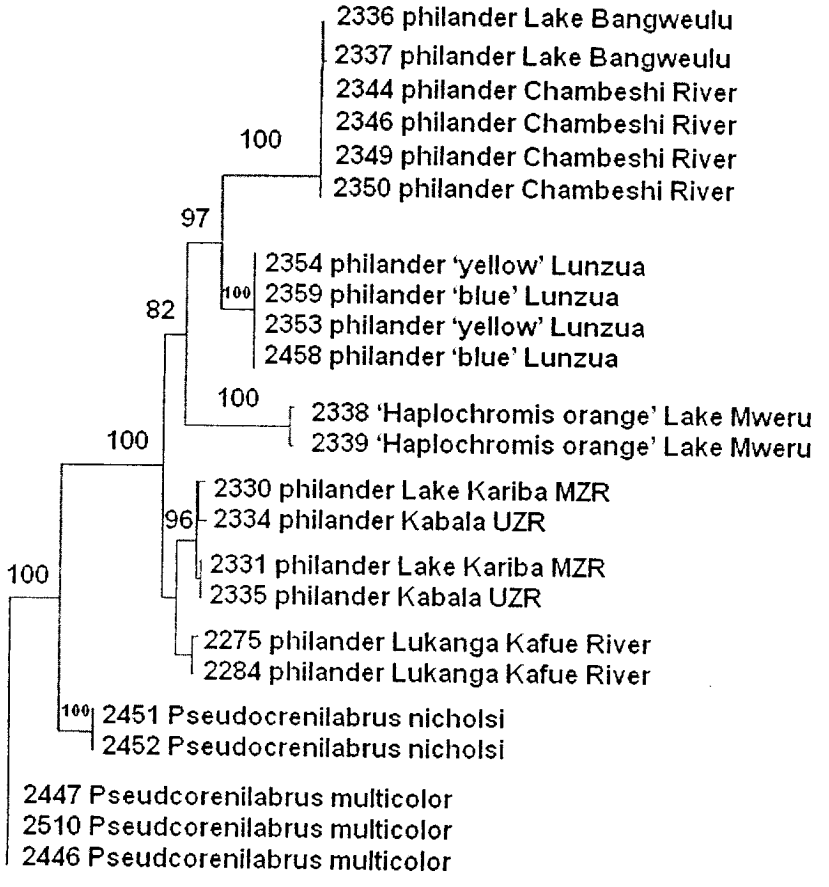


Figure 54. Neighbour joining tree [substitution model HKY+G (Hasegawa *et al.*, 1985)] of 18 taxa of the Zambian *Pseudocrenilabrus* species complex from seven localities, plus the two outgroup taxa *Pseudocrenilabrus multicolor* and *P. nicholsi*, based upon 1047 bp of the NADH2 gene. Only bootstraps larger than 50 are shown.

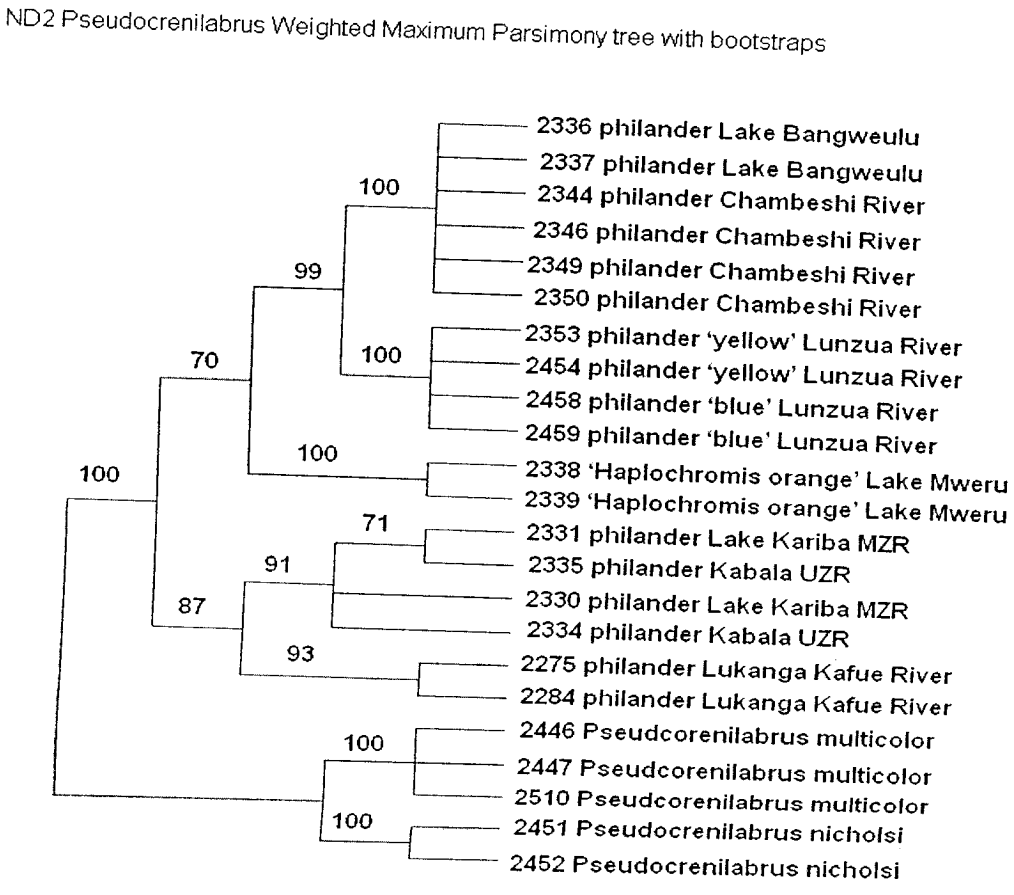


Figure 55. Strict consensus tree of two MP trees of 18 taxa of the Zambian *Pseudocrenilabrus* species complex from seven localities, plus the two outgroup taxa *Pseudocrenilabrus multicolor* and *P. nicholsi*, based upon 1047 bp of the most variable part of the NADH2 gene. Only bootstraps larger than 50 are shown.

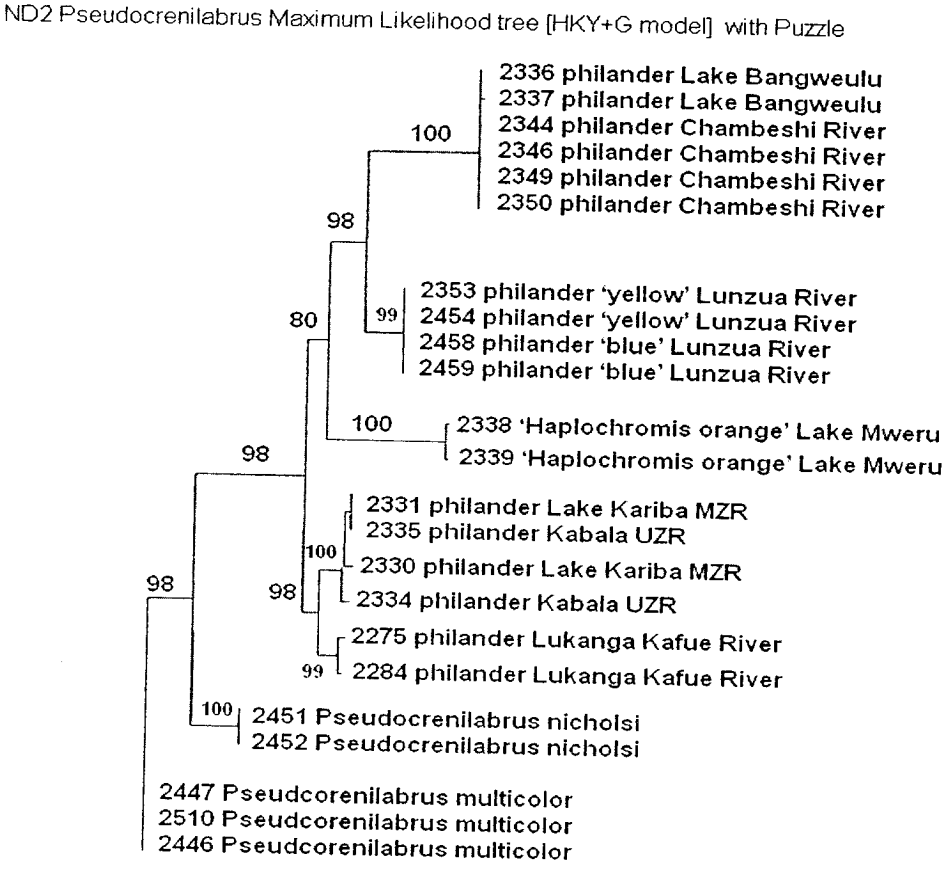


Figure 56. Maximum Likelihood tree [substitution model HKY +G (Hasegawa *et al.*, 985)] of 18 taxa of the Zambian *Pseudocrenilabrus* species complex from seven ocalities, plus the two outgroup taxa *Pseudocrenilabrus multicolor* and *P. nicholsi*, based pon 1047 bp of the NADH2 gene. Only quartet puzzling values larger than 50 are nown.

4.2.1.2.4 *Pseudocrenilabrus* ND2 Four Cluster Likelihood Mapping analysis

This analysis gives very clear biogeographic pattern but unclear phylogeographic affinities. The order of colonization is unclear. This may be due sampling bias for example the Lufubu River has not been sampled yet it is closely related to the other systems sampled. The 4-Cluster LMA should be given high weighting in interpreting the overall results. There is a strong support of bifurcation in one direction (a,b) – (c,d): Chambeshi/Bangweulu/Lake Mweru/ + Kafue/Upper Zambezi / Lunzua (figure 50).

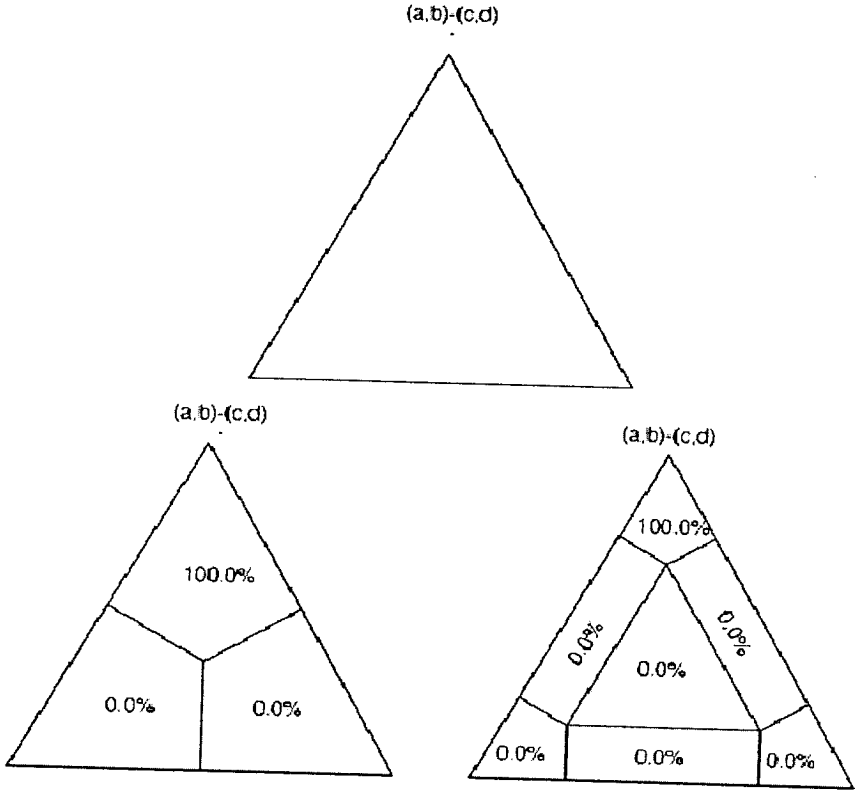


Figure 57. Four Cluster Likelihood Mapping Analysis of Zambian *Pseudocrenilabrus* species complex based on the mitochondrial NADH2 gene. The clusters are Bangweulu and Chambeshi populations (a), Lake Mweru population (b), Zambezi and Kafue populations (c). The Lunzua population forms cluster d. This analysis strongly supports the (ab) – (cd) topology.

4.2.2 Molecular phylogenetic analysis of Zambian serranochromine cichlids

4.2.2.1 Serranochromine D-Loop phylogeny

For fine scale analysis of closely related serranochromine taxa, the control region (D-Loop) was used. The D-Loop is very fast evolving and will contain relatively new branches within closely related taxa. The control region is non-protein coding and therefore has more mutations than the protein-coding genes. This region therefore gives the fine-scale phylogeny within the serranochromine clades. The most variable part of the D-Loop (Figure 59) was used in this analysis. The outgroup taxa used in this analysis were chosen based on the Tanganyika phylogeny (Figures 68 and 69).

4.2.2.1.1 Serranochromine D-Loop Neighbour Joining (NJ) tree

This tree was constructed based on the TrN model (Tamura and Nei, 1993) with a gamma distribution shape parameter (G) of 0.2340. Base frequencies were: A (0.3657), C (0.1972), G (0.1323) and T (0.3047), as depicted in Figure 58. The NJ analysis gives the first-glance phylogenetic tree of the taxa. The serranochromine species studied included *Serranochromis* (98 specimens), *Sargochromis* (43 specimens) and *Pharyngochromis* (five specimens). *Cyclopharynx fwae* (one specimen), and *Orthochromis polyacanthus* (one specimen), both from the Upper Congo, were used as out-group taxa. The tree defines five lineages, which cluster in two major clades. The first clade comprises two lineages while the second clade contains three lineages (Figure 58).

The first lineage comprises *Serranochromis stappersi* from the Mwatishi River/Lake Mweru confluence and from Mifimbo, the estuarine delta where the Luapula River flows into Lake Mweru. This lineage is supported by a bootstrap value of 86 indicating strong support of monophyly. This lineage is exclusively Congo based and is not found in the Zambezi drainage.

The second lineage is supported by a bootstrap value of 84 and comprises *Serranochromis robustus* from Mambova (Upper Zambezi), Barotse (Upper Zambezi) Chanyanya (Kafue River), Mazabuka (Kafue River) and *Serranochromis thumbergi* from Lake Bangweulu (Luapula-Congo drainage).

The third lineage mainly comprises *Serranochromis angusticeps* and *S .altus* although it also has an undescribed *Serranochromis* from Mifimbo, Lake Mweru (bootstrap value for this lineage is 84). The *S. angusticeps* belonging to this lineage are from Lakes Bangweulu and Mweru (Luapula-Congo drainage); Chambeshi (Luapula-Congo drainage); Mambova (Upper Zambezi); Barotse (Upper Zambezi); Chanyanya, Mazabuka and Lukanga (Kafue River). There is a partial separation of *S. angusticeps* of the Congo from the rest except two individual from Mambova, Upper Zambezi. The *S. altus* in this lineage are from Barotse Flood Plains and Mambova (Upper Zambezi). Incomplete lineage sorting is observed between *S. angusticeps* and *S. altus*. There is polyphyletic placement of the species of this lineage in terms of taxonomy. The undescribed species from Lake Mweru morphologically resembles *S. macrocephalus* and *S. stappersi* (Figures 28, 29 and 30).

The fourth lineage comprises *Serranochromis macrocephalus* and *Serranochromis longimanus*. The bootstrap value for this lineage is 95. The members of this lineage come from: Baroste and Mambova (Upper Zambezi); Lukanga, Chanyanya and Mazabuka (Kafue); Sinazongwe and Siavonga (Middle Zambezi). Only two specimens of *S. longimanus* were collected from Barotse (Upper Zambezi) during fieldwork.

The fifth lineage comprises *Sargochromis* species and *Pharyngochromis species*. The bootstrap value for this lineage is 87. The *Sargochromis* have not all been completely identified to species level. Those identified are *S. giardi* (Mambova and Barotse, Upper Zambezi), *S. mellandi* (Chambeshi, Bangweulu and Mweru, Congo drainage) and *S. codringtoni* (Siavonga, Middle Zambezi). A number of *Sargochromis* from the Kafue and a few of them from Upper Zambezi remain unidentified. The *Pharyngochromis acuticeps* in this lineage are from Siavonga and Sinazongwe (Lake Kariba, Middle Zambezi). The

Pharyngochromis species from Mambova (Upper Zambezi) could be different from *P. acuticeps* from Lake Kariba. There is no complete lineage sorting between the *Sargochromis* and *Pharyngochromis* genera although there is a good support for the monophyly of this lineage.

The branch between the fourth and fifth lineages is very short, suggesting a contemporaneous origin of the two lineages. Its low bootstrap value (42) indicates a very weak support for the monophyly of these two lineages.

4.2.2.1.2 Serranochromine D-Loop Maximum Parsimony (MP) tree

The strict consensus tree of two MP trees is depicted in Figure 60. This analysis is based on 358 bp of the most variable part of the control region (Figure 59). Only bootstrap values larger than 50 are shown. For this analysis, a reduced data set was used in which identical sequences were represented by only one. This was done to reduce on computer time needed to calculate the tree. The serranochromines studied were *Serranochromis* (27 specimens), *Sargochromis* (10 specimens) and *Pharyngochromis* (two specimens). *Cyclopharynx fwae* (one specimen), and *Orthochromis polyacanthus* (one specimen), were used as out-group taxa. The bootstrap was carried out using the fast stepwise addition option with 10,000 replicates. This was done because MP bootstrap needs a lot of computer time to complete. The branching order is similar to that obtained in the NJ tree but the overall bootstraps for the MP tree are lower. Again, the five lineages are grouped into two major clades, which is consistent with the NJ tree.

4.2.2.1.3 Serranochromine D-Loop Maximum Likelihood (ML) tree

This tree was constructed based on the TrN model (Tamura and Nei, 1993) with a gamma distribution shape parameter (G) of 0.5533. The proportion of invariable sites (I) was 0.3937. Base frequencies were: A (0.3608), C (0.1989), G (0.1341) and T (0.3062). See Figure 61 for the ML tree. This analysis was carried out with the same number of individuals as in the case of the MP tree but instead of bootstrap analysis, the puzzle

analysis was carried out due to computing time constraints. This analysis also supports five lineages. In contrast to NJ and MP, *Serranochromis stappersi* occupies the most ancestral branch.

4.2.2.1.4 Serranochromine D-Loop Four-Cluster Likelihood Mapping Analysis

The taxa were grouped into four clades in which there are three possible topologies. The likelihood support for each topology was then calculated according to the ML tree. The clades are defined as follows: a = *Serranochromis stappersi* + *Serranochromis robustus;* b = *S. angusticeps* + *S. altus* + *Serranochromis 'sp elongate'* ; c = *S. macrocephalus* and d = *Sargochromis* + *Pharyngochromis.* Only one topology is supported by this analysis i.e. (a,b)-(c,d).See Figure 62.

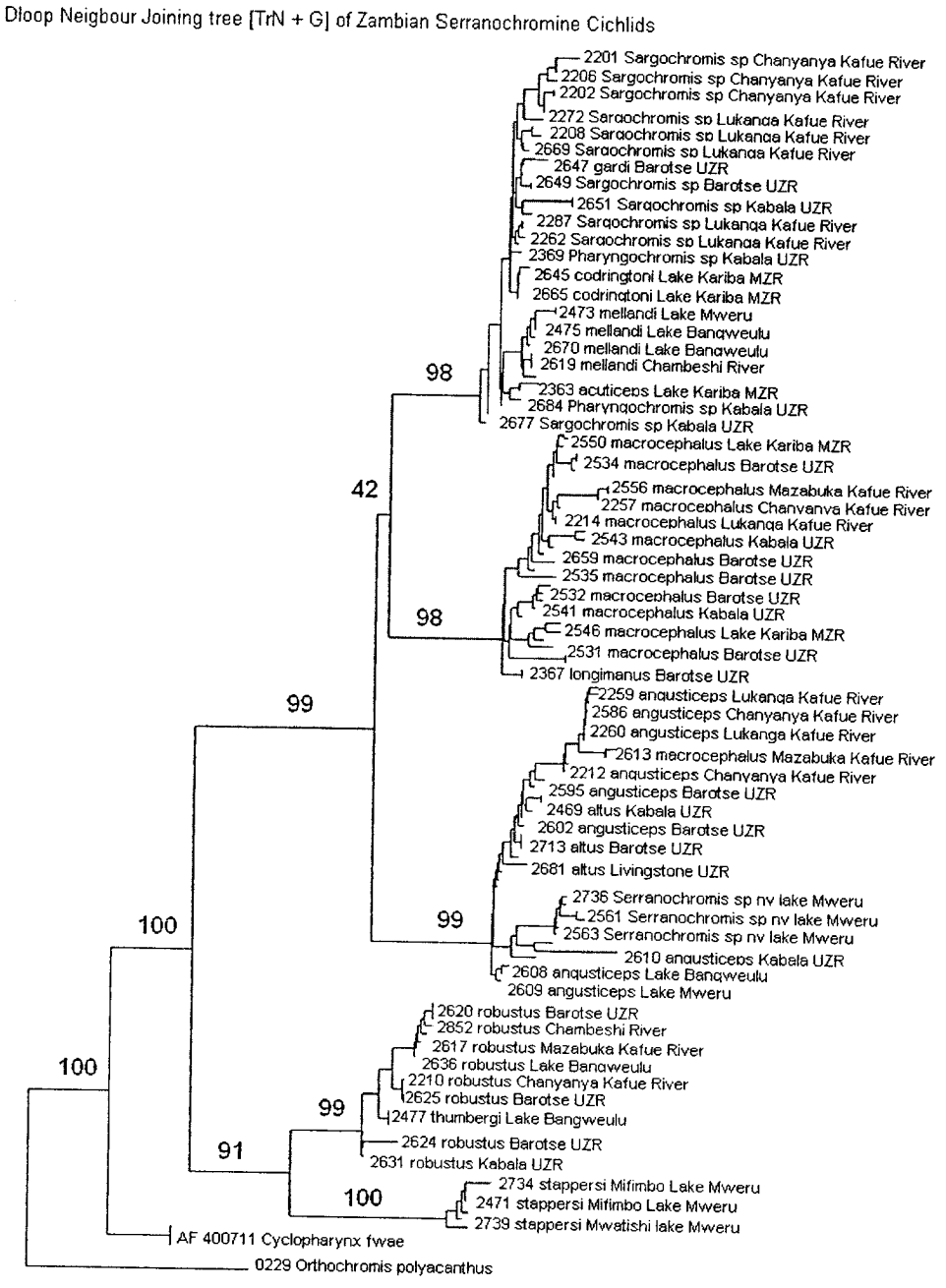


Figure 58. The Neigbour Joining tree [TrN + G] of Zambian Serranochromine Cichlids based on the most variable part (Dloop) of control region of the mitochondrial genome. Only bootstrap values above 40 are shown. *Serranochromis* 'sp nv' is the same group referred to as *Serranochromis* 'sp elongate' elsewhere in this report.

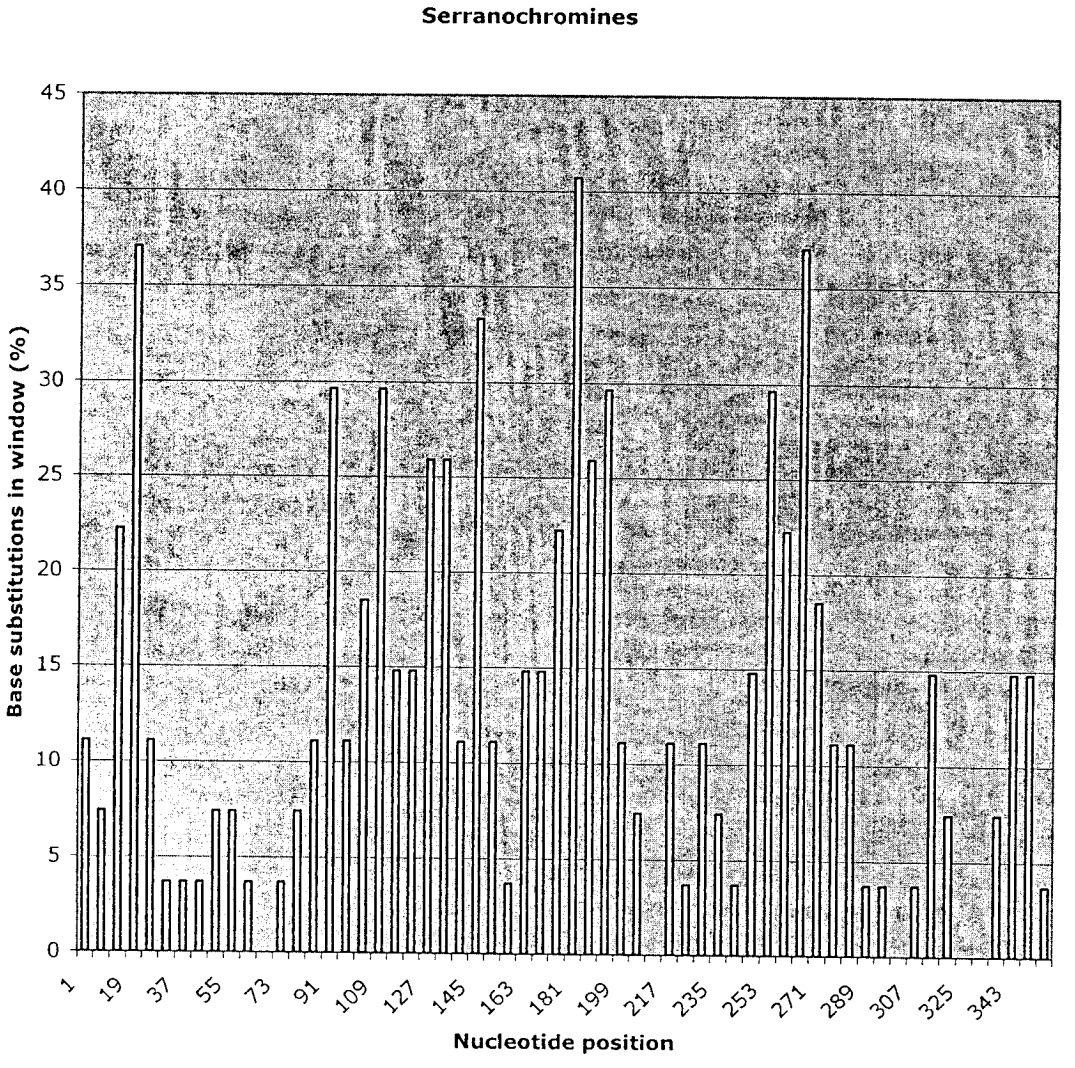


Figure 59. Sliding window test based on a 358 bp segment of the most variable part of the mitochondrial control region. Three regions were defined as showing low variability (Low Var) if the overall base substitution was less than ten percent (<10%), high variability (High Var) if the overall base substitution was between ten and twenty percent (10-20%) and hyper variability (Hyper Var) if the overall base substitution was more than twenty percent (>20%).

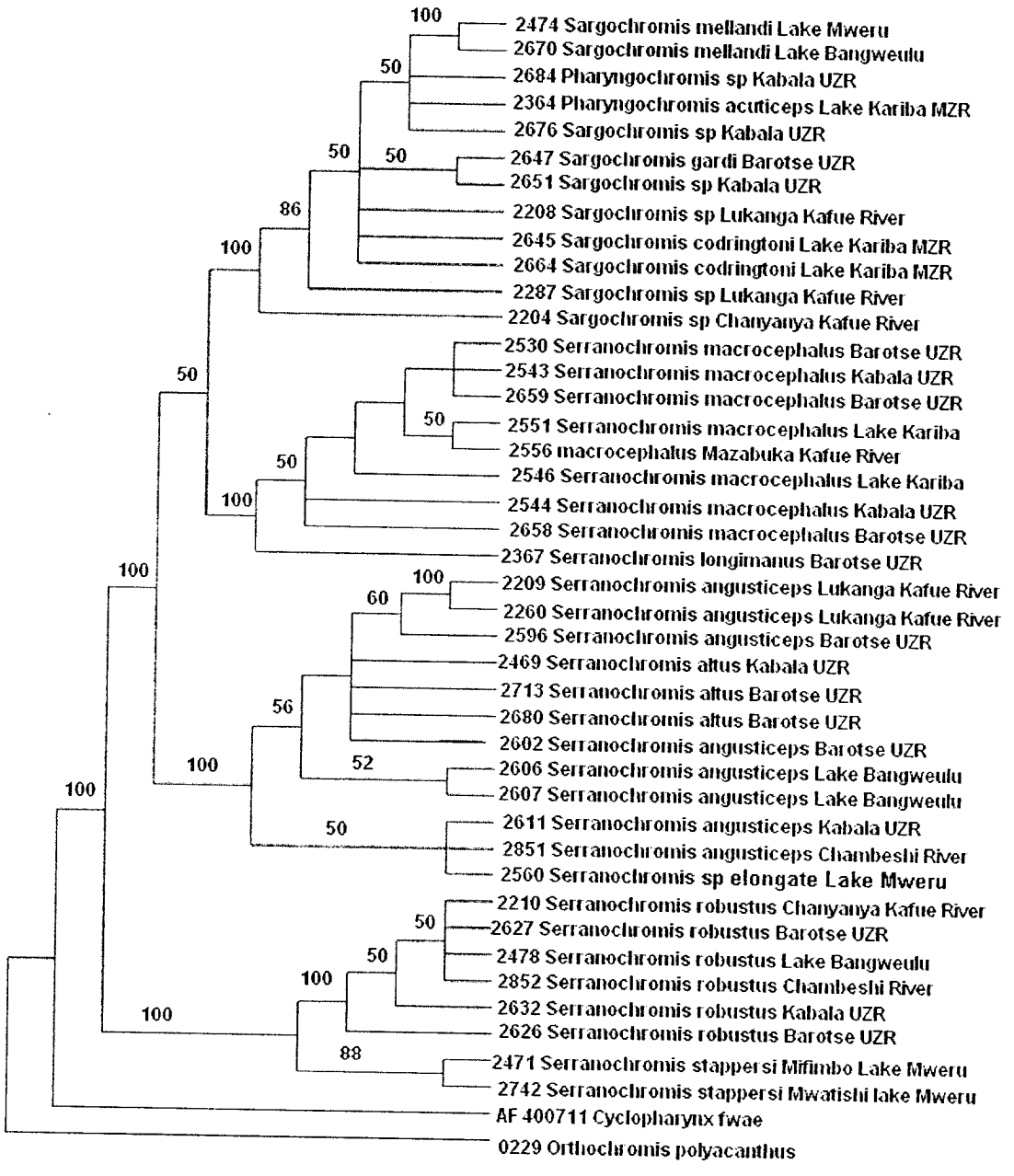


Figure 60. Strict consensus tree of two MP trees of the Zambian riverine serranochromine cichlid species from seven localities, plus the two outgroup taxa *Cyclopharynx fwae* and *Orthochromis polyacanthus,* based upon 358 bp of the most variable part of the mitochondrial control region. Only bootstraps larger than 50 are shown.

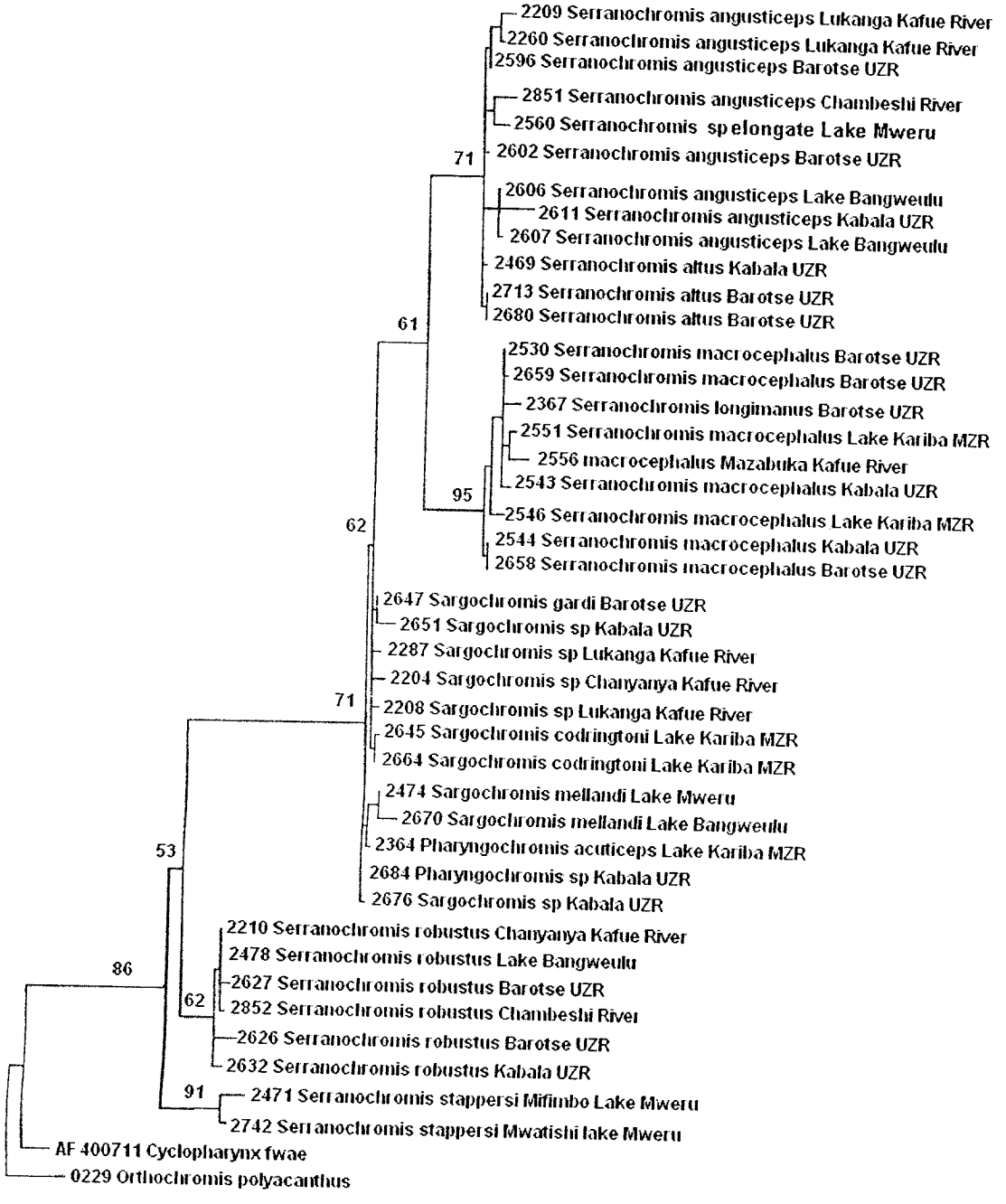


Figure 61. Maximum Likelihood tree [substitution model TrN+I +G (Tamura and Nei, 1993)], of the Zambian riverine serranochromine cichlid species from seven localities, plus the two outgroup taxa, *Cyclopharynx fwae* and *Orthochromis polyacanthus,* based upon 358 bp of the most variable part of the mitochondrial control region. Only puzzle values larger than 50 are shown.

Serranochromine D-Loop Four-Cluster Likelihood Mapping Analysis

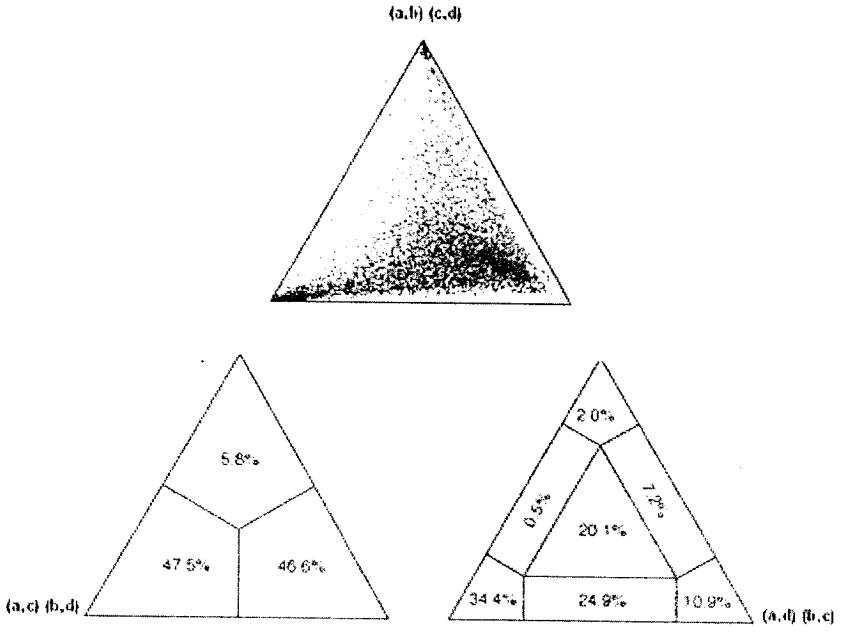


Figure 62. Serranochromine DLoop 4 Cluster Likelihood mapping Analyis based on the most variable part of the mitochondrial control region. This analysis supports two alternative topologies, i.e. (a,c)-(b,d) and the (a,d)- (b,c). The (a,c) – (bd) topology is the more favoured one with a value of 34,4% compared to the (a,d)- (b,c) which has a value of 10,9%.

4.2.2.1.5 Serranochromine linearized tree

This tree was based on the TrN +G model (Tamura and Nei, 1993). The analysis shows that there may be two major cladogenesis events: one old split and one young split. Two major groups; *S. robustus* + *S. stappersi* and the *Sargochromis/Pharyngochromis* + *Serranochromis angusticeps/ S. altus* + *Serranochromis* 'sp elongate' are evident (Figure 6).

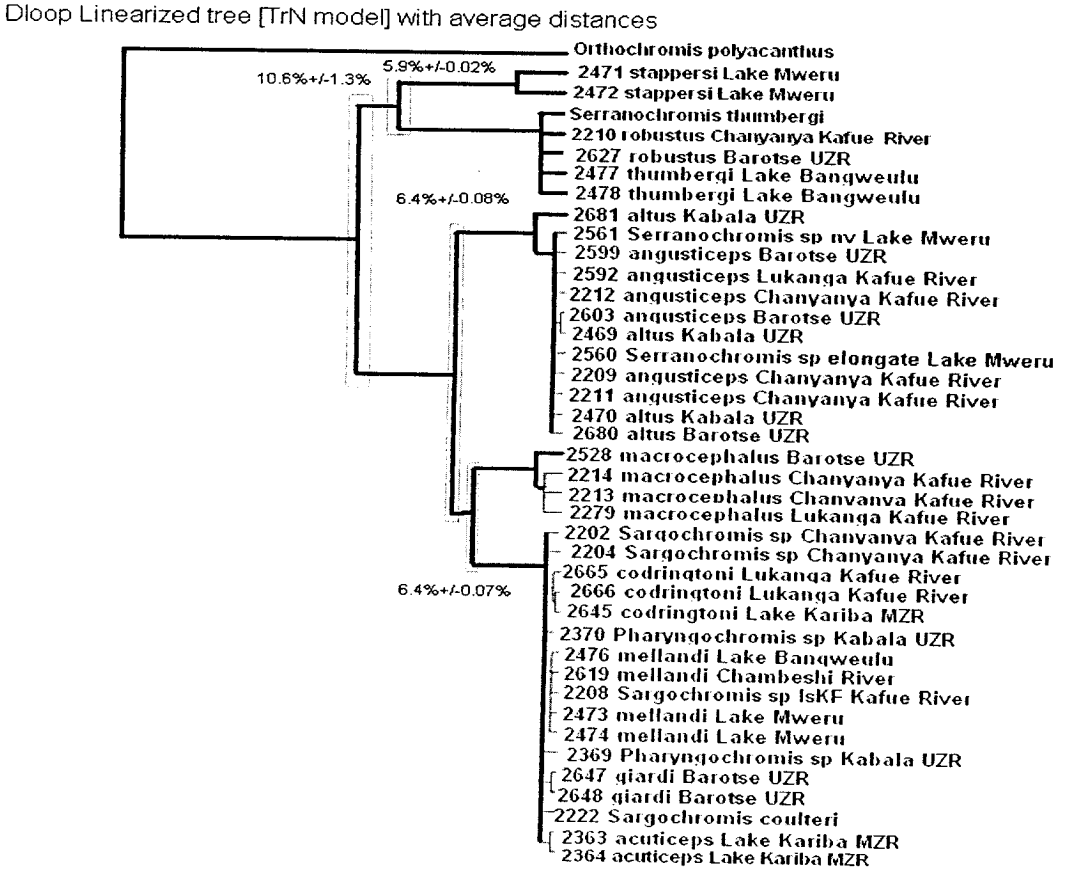


Figure 63. Linearized tree based on a 358 bp segment of the most variable part of the mitochondrial control region. The linearized tree was compiled using the computer program LINTRE (Takezaki *et al.*, 1995) after performing a branch length test (Takezaki *et al.*, 1995) to test for differences in base substitution rates, using the substitution model TrN (Tamura and Nei, 1993). Grey rectangles refer to major diversification events. The observed mean sequence divergences using the substitution model K2P (Kimura, 1980) are depicted for the corresponding diversification events. The identities of the specimens used and the localities from where they were collected are indicated.

Two values were obtained for average distances, with the first value estimated from the TrN +G model to get the scale for the tree. The second value was estimated using the K2P model (Kimura, 1980) to be able to compare the present results with those from literature. There is an interesting and corroborative result (Figure 63). This tree was based on the TrN +G model but the average distances used were based on the K2P model. There is one old split from which the *Serranochromis stappersi* and *Serranochromis robustus* arose parallel to the rest of the serranochromines. The more recent split resulted into a radiation of *Serranochromis*, *Sargochromis* and *Pharyngochromis*. Two major groups are identifiable: the *S. robustus/S. thumbergi / S. stappersi* group and the *Sargochromis/Pharyngochromis / Serranochromis angusticeps/ S. altus /Serranochromis* 'sp elongate' group. There is a very recent divergence within the five lineages. The age of each of the five lineages is similar.

4.2.2.2 Serranochromine ND2 phylogeny

For deep branches in the phylogeny, the ND2 gene is used because it is protein-coding and therefore better conserved than the non-coding D-Loop. It evolves more slowly than the Dloop because it cannot survive too many mutations. Therefore, the ND2 gene was used to give the overall phylogeny of the clades. For this analysis, the same strategy as that used for the Dloop analysis was followed. The outgroup taxa used were chosen based on the Tanganyika phylogeny (Salzburger *et al.*, 2002; Figures 68 and 69).

4.2.2.2.1 Serranochromine ND2 Neighbour Joining (NJ) tree

This tree was constructed based on the TrN model (Tamura and Nei, 1993) with a gamma distribution shape parameter (G) of 0.2781. Base frequencies were; A (0.2614), C (0.3470), G (0.1255) and T (0.2661). See Figure 64. The Serranochromine genera studied were *Serranochromis* (22 specimens), *Sargochromis* (12 specimens) and *Pharyngochromis* (four specimens). *Cyclopharynx fwae* (one specimen), and *Orthochromis polyacanthus* (one specimen), were used as out-group taxa. This tree highly supports four clades with good bootstraps (Figure 64). *Serranochromis stappersi* is placed as the most ancestral branch in this analysis.

4.2.2.2.2 Serranochromine ND2 Maximum Parsimony tree

The strict consensus tree of the two most parsimonious trees is depicted in Figure 65. This analysis was based on 1047 bp of the ND2 gene of the mitochondrion. Only bootstrap values larger than 50 are shown as in Figure 65. The number of evolutionary steps, consistency index and rescaled consistency index are not shown in this figure. *Serranochromis stappersi* is placed as the most ancestral branch in this analysis like in the NJ tree above.

4.2.2.2.3 Serranochromine ND2 Maximum Likelihood tree

This tree was constructed based on the TrN model (Tamura and Nei, 1993) with a gamma distribution shape parameter (G) of 0.2781. Base frequencies were; A (0.2614), C (0.3740), G (0.1255) and T (0.2661). This tree highly supports the four clades with good bootstraps. See Figure 66. *Serranochromis stappersi* is placed as the most ancestral branch.

4.2.2.2.4 ND2 serranochromine four-cluster Likelihood Mapping analysis

The four clusters were identified as a = *Serranochromis stappersi*, *S. robustus* and *S. thumbergi*, cluster b = *Serranochromis altus*, *S. angusticeps* and *Serranochromis* 'sp elongate', cluster c = *Serranochromis macrocephalus*. *Sargochromis* and *Pharyngochromis* were placed in cluster d. This analysis strongly supports the (ab) – (cd) topology. This suggests a sister relationship between *Serranochromis stappersi/S. robustus* and *Serranochromis altus/ S. angusticeps*. A similar sister relationship is suggested between *Serranochromis macrocephalus* and *Sargochromis/Pharyngochromis*. See Figure 67.

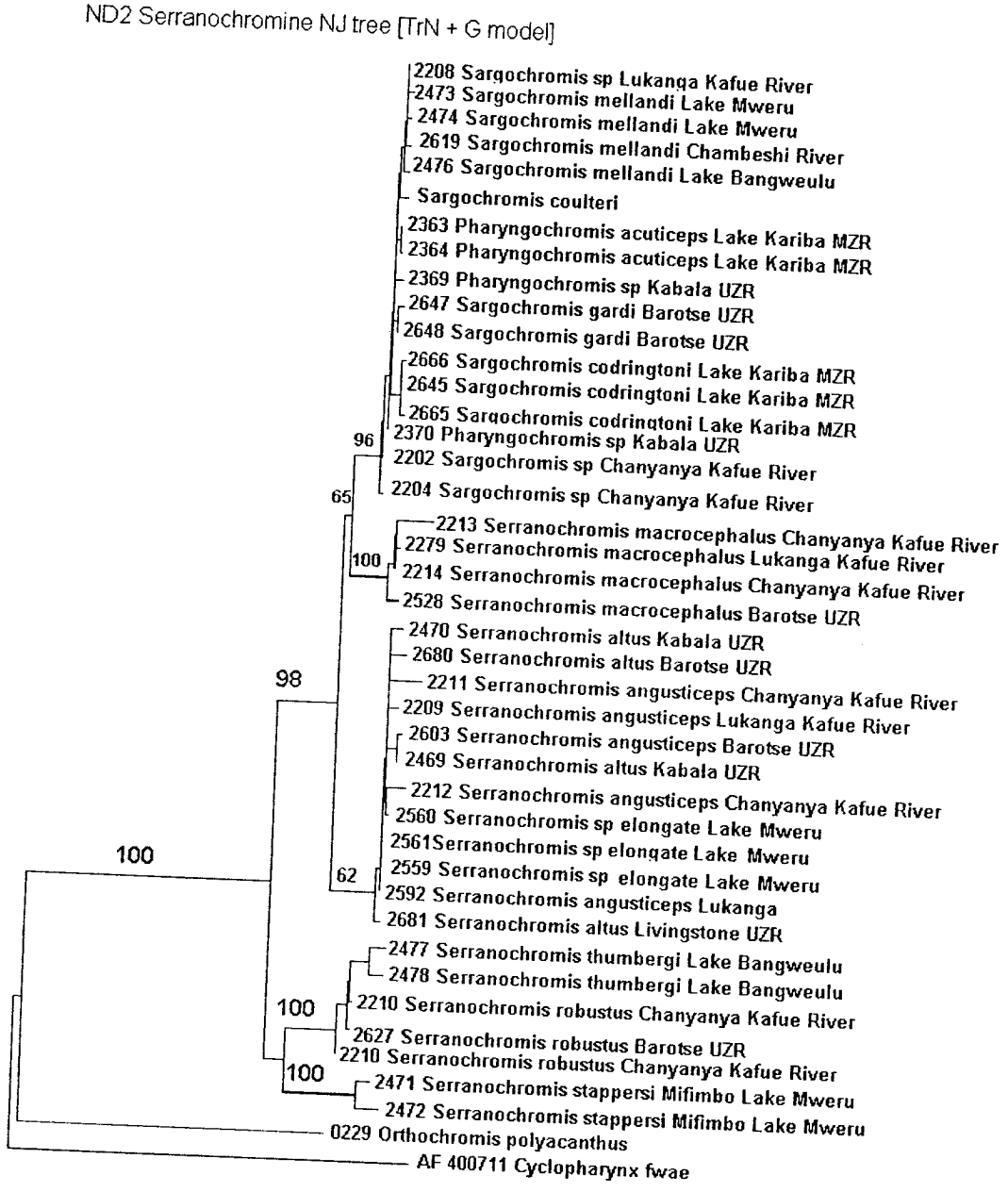


Figure 64. The Neigbour Joining tree [TrN + G] of Zambian serranochromine cichlids, based on the ND2 gene is shown above. Only bootstap values above 50 are shown.

ND2 Serranochromine Maximum Parsimony tree (strict) with bootstrap

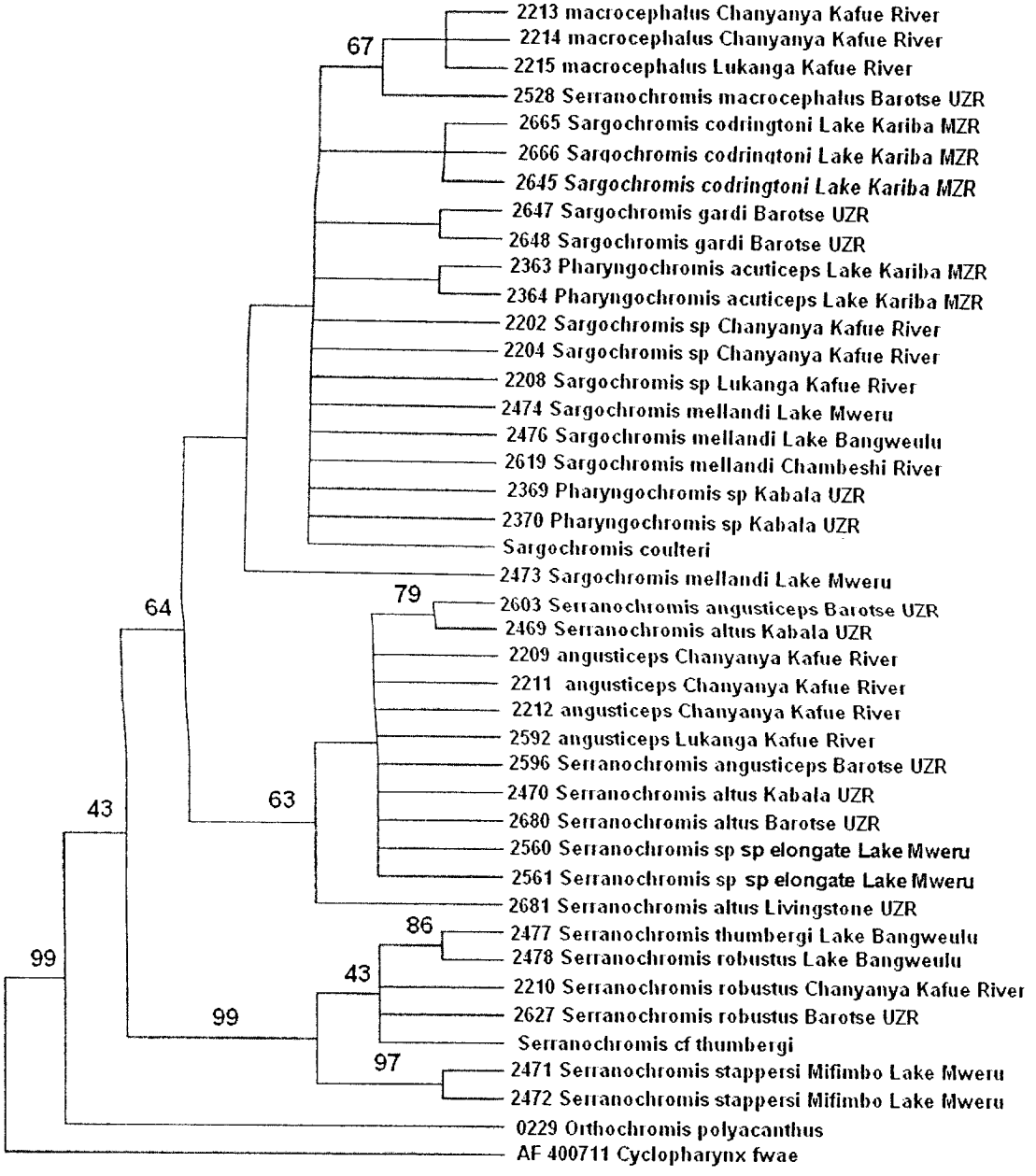


Figure 65. Strict consensus tree of two MP trees of the Zambian riverine serranochromine cichlid species from seven localities, plus the two outgroup taxa *Cyclopharynx fwae* and *Orthochromis polyacanthus,* based upon ND2 gene of the mitochondrial DNA. Only bootstrap values larger than 50 are shown.

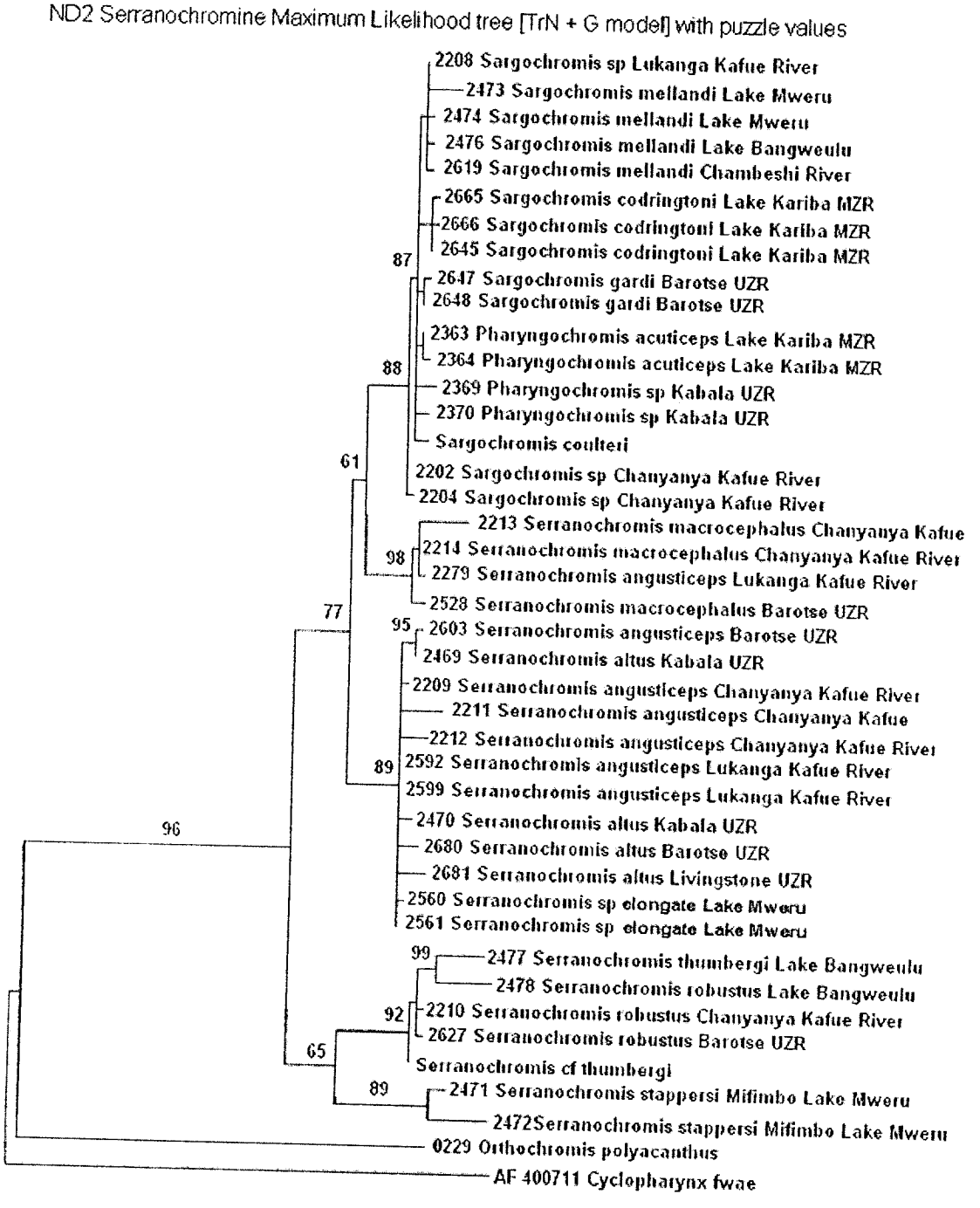


Figure 66. Maximum Likelihood tree [substitution model TrN+I +G (Tamura and Nei, 1993)] of the Zambian riverine serranochromine cichlid species from seven localities, plus the two outgroup taxa, *Cyclopharynx fwae* and *Orthochromis polyacanthus,* based upon ND2 gene of the mitochondrial DNA. Only puzzle values larger than 50 are shown.

ND2 Serranochromine Four Cluster Likelihood Mapping Analysis

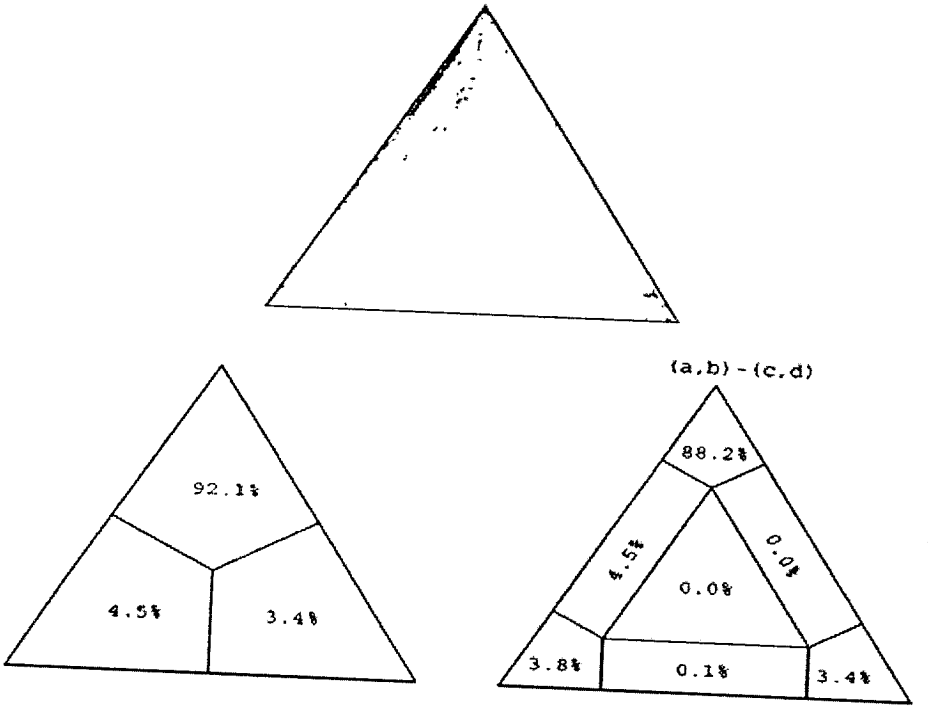


Figure 67. Four-cluster Likelihood Mapping analysis of Zambian serranochromine cichlids based on the ND2 gene. The clusters are *Serranochromis stappersi, S. robustus* and *S. thumbergi* (a), *Serranochromis altus* and *S. angusticeps* (b), *Serranochromis macrocephalus* (c). *Sargochromis* and *Pharyngochromis* form cluster d. This analysis strongly supports the (ab) – (cd) topology.

4.3 Phylogenetic placement of the *Pseudcrenilabrus* and serranochromine species

The *Pseudocrenilabrus* have a sister group relationship with other members of the Haplochromini and with the Tropheini while the serranochromines have a sister relationship with *Orthochromis* within the Tanganyika radiation (Figures 68 and 69). The *Pseudcrenilabrus* and serranochromine species both belong to the Haplochromimi (Salzburger, 2002).

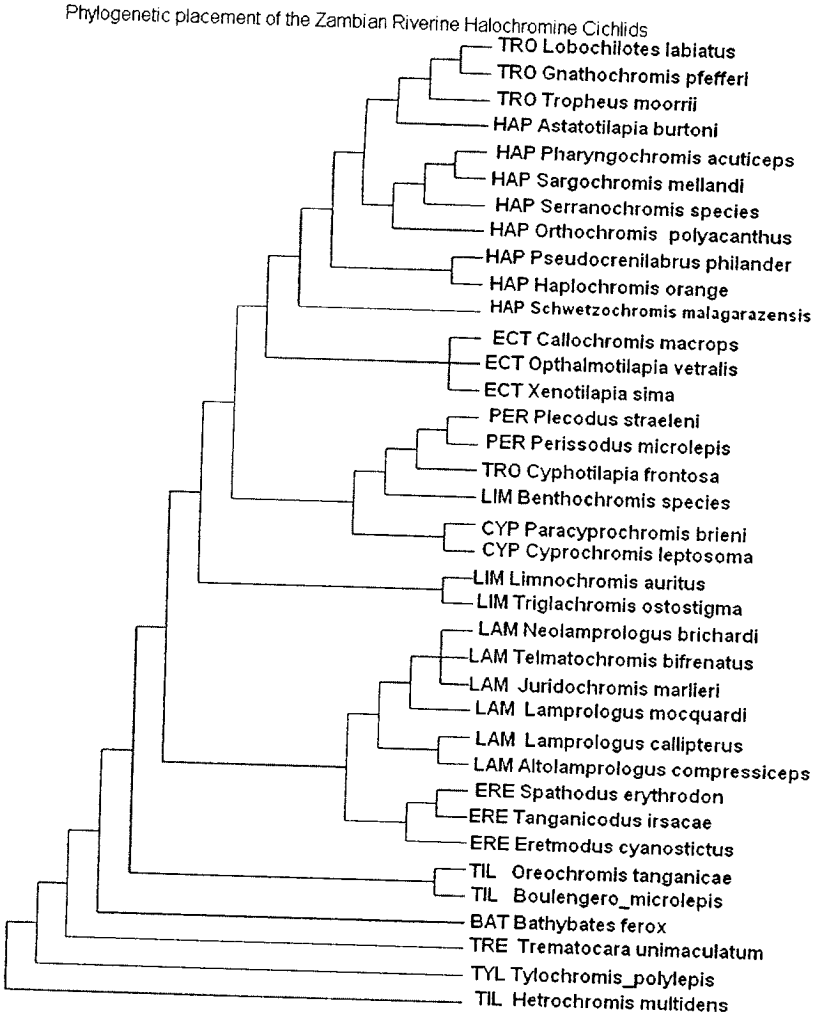


Figure 68. The Phylogenetic placement of the *Pseudcrenilabrus* and Serranchromis species shown above is based on an Unweighted Maximum Parsimony tree representing 9 distinct tribes [ERE, Eretmodini; LAM, Lamprologini; TYL, Tilapiini; LIM, Limnochromini; CYP, Cyprochromini; HAP, Haplochromini, PER, Perissodini; ECT, Ectodini; TRO, Tropheini.] This assignment to tribes follows Salzburger et al. (2002) based on 1047 bp of the NDH 2 gene and 402 bp of the cytochrome *b* gene. *Bathybates ferox, Trematocara unimaculatum, Tylochromis polylepis* and *Heterochromis multidens* were used as outgroup taxa. The unusual position of the specimen of *Benthochromis species* could be due to incomplete resolution of the tree or misidentification.

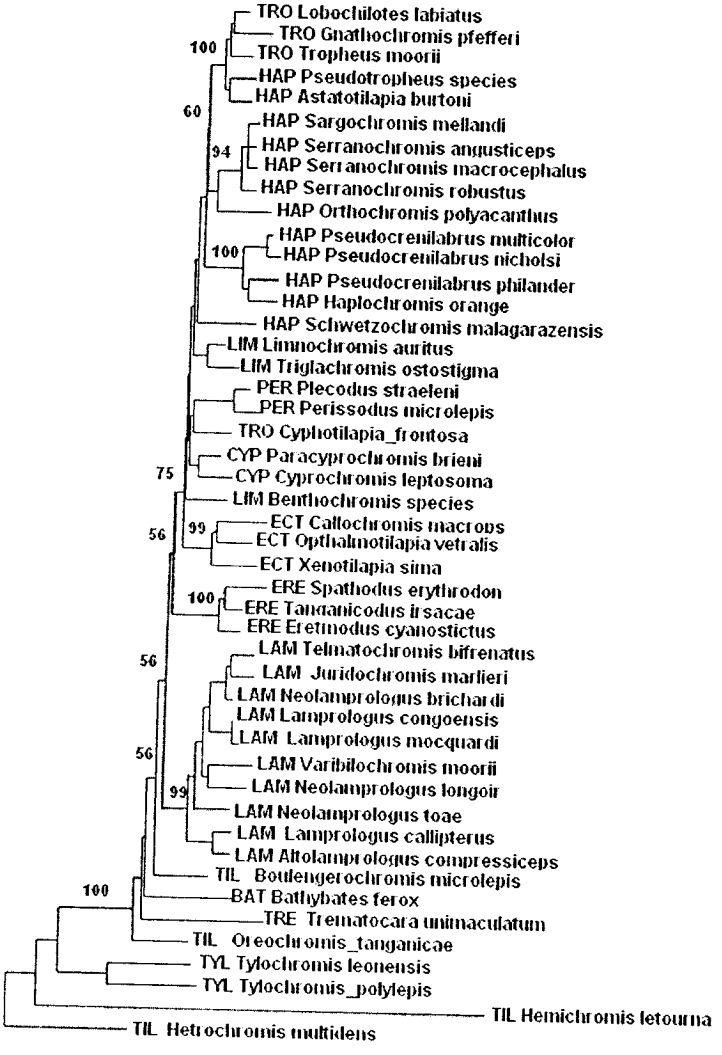


Figure 69. The Phylogenetic placement of the *Pseudcrenilabrus* and Serranchromine species shown above is based on the Neighbour Joining tree representing 9 distinct tribes [ERE, Eretmodini; LAM, Lamprologini; TYL, Tilapiini; LIM, Limnochromini; CYP, Cyprochromini; HAP, Haplochromini, PER, Perissodini; ECT, Ectodini; TRO, Tropheini. This assignment to tribes follows Salzburger *et al.* (2002) based on 1047 bp of the NDH 2 gene and 402 bp of the cytochrome b gene using the substitution model TrN + I + G (Tamura and Nei, 1993). *Bathybates ferox, Trematocara unimaculatum, Hemichromis letourna* and *Heterochromis multidens* were used as outgroup taxa. The specimen of *Benthochromis species* has a sister relationship with the tribe Cyprochromimi. This unusual result could be due incomplete resolution of the tree or misidentification.

4.4. Combined morphological and molecular biological analyses of the riverine haplochromine cichlids

In carefully planned systematic and phylogenetic studies, the results from morphological data should be similar to those obtained from molecular data based on the same sample (Patterson *et al.*, 1993; Lippitsch, 1997; Ruber *et al.*, 1999; Farias *et al.*, 2000; Kuusipalo, 2000; Kai *et al.*, 2002 and Pouyaud *et al.*, 2003). In this study, phylogenetic analysis based on morphological data was not carried out. Only PCA's were carried out on morphological data. All phylogenetic analyses were based on mitochondrial DNA sequences. More data were generated from morphological analysis than from molecular analysis because it was easier to generate morphological data. The generation of molecular data was constrained by the cost of reagents, low success rate in PCR, CTR and automatic sequencing steps and time required to repeat unsuccessful runs.

The evolutionary trees reconstructed in this study primarily reflect the evolution of mitochondrial DNA while the morphological characters were used mainly for taxonomic review purposes. The morphometric and meristic characters used for the taxonomic review of the riverine haplochromine species were based on characters of proven taxonomic content (Barel *et al.*, 1977). The morphological characters, which are considered to be of evolutionary importance in cichlids, are: jaw shape and size; head shape and size and breeding male colour patterns. The morphology of the head and jaws is very important in the trophic specialization of cichlids (Kornfield and Smith, 2000). The male breeding colours, though not used in this study, are also important diagnostic characters for delineating species. Colour patterns are thought to be driven by sexual selection, since females also base their choice of mates on male colours (Turner and Burrows, 1995; Higashi *et. al,* 1999; Seehausen *et al.*, 1999). Among the morphological characters used in this study, the jaw and head dimensions were contributing a lot to the taxonomic differences among the taxa. Therefore, the PCA results helped to evaluate the distinctness of species and populations and also contained evolutionary information about the riverine haplochromines.

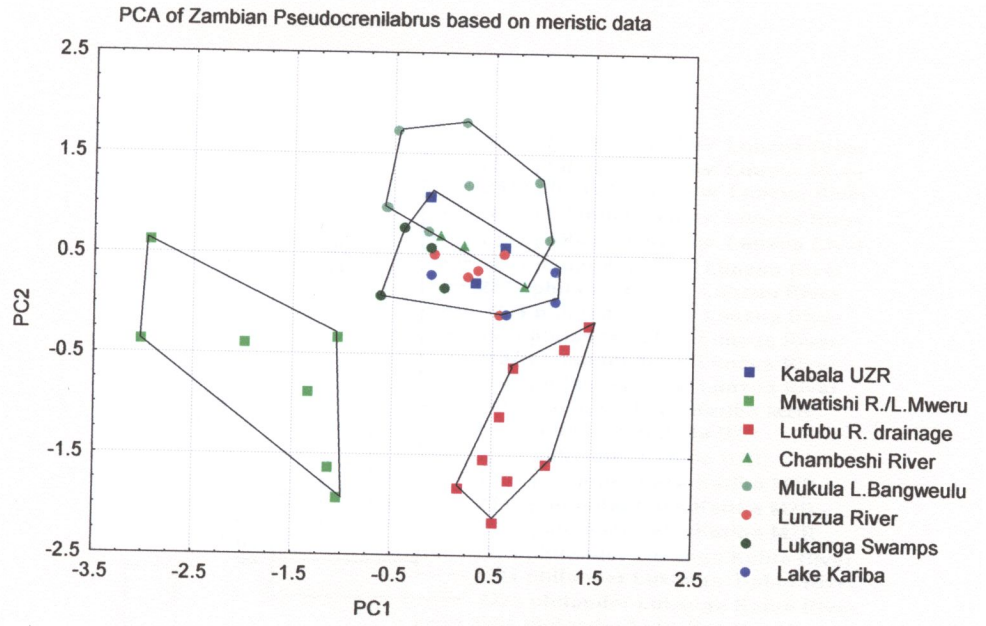


Figure 70. PCA of the meristics of the Zambian *Pseudocrenilabrus* species complex. This is a PC1 x PC2 plot, N = 43. The population from the Mwatishi River/Lake Mweru confluence separates out on PC1. The population from the Lufubu River drainage separates from the populations from the Lunzua River, Luapula-Chambeshi and the Zambezi-Kafue drainages on PC2. See Chapter 3, Figure 15 for details of this PCA.

Figure 71. Strict consensus tree of two MP trees [246 steps; CI excluding uninformative sites, 0.82; retention index (RI), 0.96; rescaled consistency index (RC), 0.79] the NJ tree and the ML tree, of 38 taxa of the Zambian *Pseudocrenilabrus* species complex from seven localities, plus the two outgroup taxa *Pseudocrenilabrus multicolor* and *P. nicholsi*, based upon 358 bp of the most variable part of the control region. Bootstrap values obtained from neighbour joining are shown above the branches, while numbers in the middle represent parsimony bootstrp values. Quartet puzzling values from the ML tree are depicted below the branches. Only bootstrap and quartet puzzling values larger than 50 are shown.

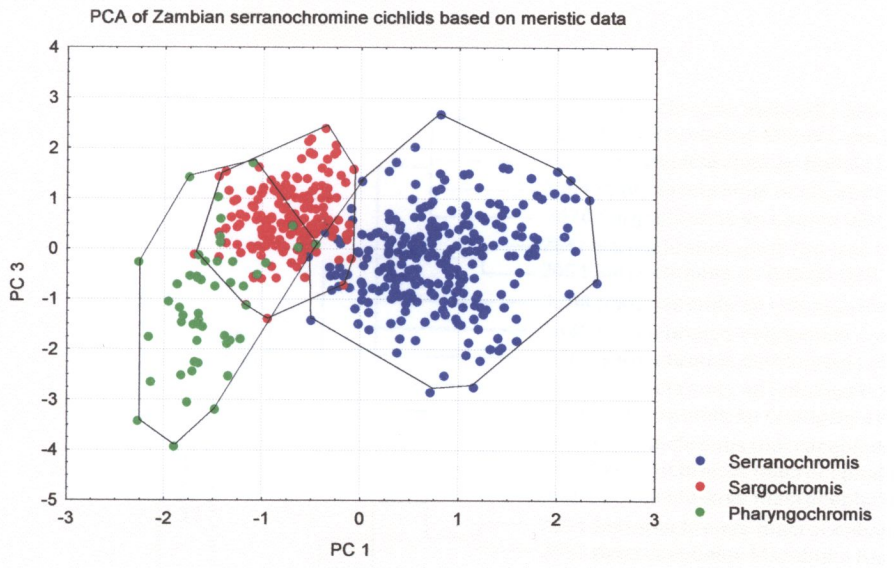


Figure 72. PCA based on meristic data, of the Zambian riverine serranochromines. This plot was between PC1 and PC3. N = 531. There is a very clear separation between *Serranochromis* and *Sargochromis* on PC1, similar to that in Fig 21a. There is also a separation with some overlap between *Sargochromis* and *Pharyngochromis* on PC 3. See Figure 27b for details of this PCA.

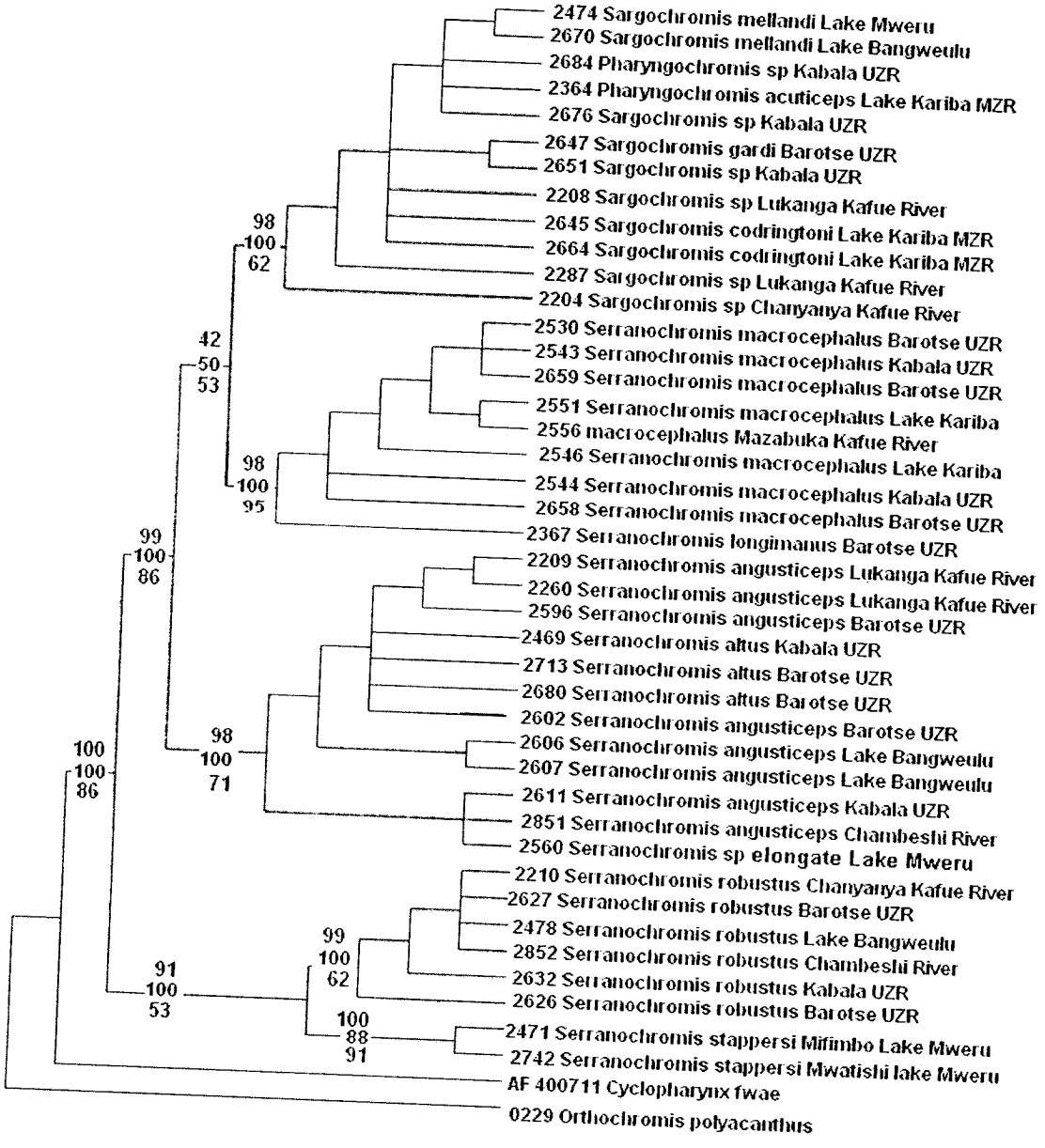


Figure 73. Strict consensus tree of two MP trees, the NJ tree and the ML tree, of the Zambian riverine serranochromine cichlid species, plus the two outgroup taxa *Cyclopharynx fwae* and *Orthochromis polyacanthus*, based upon 358 bp of the most variable part of the control region. Bootstrap values obtained from neighbour joining are shown above the branches, while numbers in the middle represent parsimony bootstrap values. Quartet puzzling values from the ML tree are depicted below the branches. Only bootstrap and quartet puzzling values larger than 50 are shown.

4.4.1. Haplochromines

It was possible to determine the phylogenetic placement of the riverine haplochromines within the Tanganyika radiation (Figures 68 and 69) using mitochondrial DNA sequence data (of Lake Tanganyika cichlids) from GENBANK (Salzburger *et al.*, 2002) together with the sequences generated from the specimens sampled from rivers during this study. However, it was not possible to carry out parallel morphological analyses on each and every specimen because it was not practically easy to locate the corresponding whole-fish samples from Lake Tanganyika.

4.4.2. *Pseudocrenilabrus*

The results from both morphological and molecular data suggest that the populations from the Tanganyika drainage system (Lunzua and /or Lufubu), the Bangweulu – Chambeshi, Lake Mweru and from the Zambezi-Kafue system belong to four different clades. It was possible to include the population from the Lufubu River only in the morphological analysis (Figure 70) but not in the molecular analysis (Figure 71) because the Lufubu specimens used were only available as formalin-preserved material from the Africa Museum at Tervuren, Belgium. Unfortunately, these specimens are not suitable for molecular (DNA) analysis.

4.4.3. Serranochromines

The results from both morphological data and molecular data suggest a close relationship between the following species-pairs: *Serranochromis angusticeps* and *S. altus*; *Serranochromis macrocephalus and S. longimanus*; *Serranochromis robustus* and *S. thumbergi* as well as between *Sargochromis* and *Pharyngochromis* species. The morphological results identify one undescribed species ('*Serranochromis* 'sp elongate') as being closely related to *Serranochromis stappersi* and *S. macrocephalus* (Figure 72) while the molecular analyses suggest the same unidentified species (*Serranochromis sp. novo* as being most closely related to the *Serranochromis angusticeps* and *S. altus*

(Figures 73). Both methods indicate that the *Sargochromis and Pharyngochromis* species need to be studied in more detail using more representatives in order to establish their exact taxonomic status and a good phylogeny.

4.5. Preliminary identification key of Zambian riverine haplochromine cichlids.

This key is based on Skelton (2001) and modified using results of the morphological character analysis from this study.

Riverine cichlid genera

1a. Teeth fine, closely set, gill rakers simple, juvenile with distinct black spot on soft-ray part of the dorsal fin (tilapia spot) .....(2)

1b.Teeth coarse, gill rakers stout, spatulate or branched, juveniles without black spot on dorsal.....(3)

2a. Adults with tilapia spot, less than 14 gill rakers on lower limb of first gill arch, substrate spawners and guarders .....***Tilapia***

2b. Adults without tilapia spot, more than 14 gill rakers on lower limb of first gill arch, mouth brooders ..... ***Oreochromis***

3a. Chest scales very small, abruptly reduced in size .....(4)

3b. Chest scales gradually reduced in size .....(5)

4a. Caudal fin rounded, males without anal fin spots or ocelli, more than 9 posterior lateral line scales separated from dorsal base by only $1^1/_2$ scales ..... ***Orthochromis***

4b. Caudal fin truncate, males with prominent anal fin ocelli, less than 9 posterior lateral line scales separated from dorsal base by only $1^1/_2$ scales ..... ***Thoracochromis***

5a. Caudal fin rounded, anal fin of males with orange tip and the maculae (small spots) in the dorsal and caudal fin, no true egg spots or ocelli, but coloured horizontal pattern in the form of an "egg streak" in the ventro posterior corner of the anal fin. .....***Pseudocrenilabrus*** (10)

5b. Caudal fin truncate, anal fin of males without orange tip, with or without egg spots or ocelli .....(6)

6a. Body with 4-5 pronounced black vertical bars, anal fin of males without egg spots or ocelli..... ***Hemichromis***

6b. Body without 4-5 pronounced black vertical bars, anal fin of males with egg spots or ocelli.....(7)

7a. Single scale row between posterior orbital margin (orbit) and vertical limb of preopercular (preopercular groove), pharyngeal bones robust, median pharyngeal teeth stout, molariform.....(8)

7b. Two or more rows of scales between posterior orbital margin (orbit) and vertical limb of preopercular (preopercular groove), pharyngeal bones slender, median pharyngeal teeth moderately large, not molariform.....(9)

8a. Last 2 or 3 pored scales in the upper lateral line separated from the dorsal fin base by not less than 2 scales of approximately equal size..... ***Sargochromis*** **(11)**

8b. Last 5 to 7 pored scales in the upper lateral line series separated from the dorsal fin base by one large and one small scale..... ***Pharyngochromis*** **(20)**

9a. Mouth very large, when closed maxilla extends to or beyond vertical through anterior orbit, teeth unicuspid ..... ***Serranochromis*****(15)**

9b. Mouth moderately large, when closed maxilla not extending to vertical through anterior orbit, teeth bi-unicuspid ..... ***Chetia***

Pseudorenilabrus **species**

10a. Lake Mweru ..... ***'Haplochromis orange'*** *

10b. Zambezi, Kafue, Chambeshi and Lake Bangweulu .... ***Pseudocrenilabrus philander***

10c. Lufubu River drainage.....***Pseudocrenilabrus*** **'sp lufubu'***

10d. Lunzua River drainage..... ***Pseudocrenilabrus*** **'sp lunzua'***

Sargochromis **species ***

11a. Head and predorsal profile straight or gently convex, mouth inclined less than 40 degrees .....(12)

11b. Head and predorsal profile straight or concave, mouth inclined 45-55 degrees ..... *Sargochromis carlottae*

12a. Preorbital depth 25-29% head length .....*S. greenwoodi*

12b. Preorbital depth 16-24% head length .....(13)

13a. Deep body, found in the Zambezi drainage .....(14)

13b. Body more slender, restricted to the Luapula drainage .....*S.mellandi*

14a. Head profile steep <40 degrees .....*S. giardi*

14b. Head profile >40 degrees ..... *S. codrongtoni* **(Zambezi)**, *S. mortimeri* **(Kafue)**

***Serranochromis* species**

15a. Length of pectoral fin 33-40 % SL ..... *S. longimanus*

15b. Length of pectoral fin >33 % SL .....(16)

16a. Lateral line scales 39-41, dorsal spines 17-18 .....*S. thumbergi*

16b. Lateral line scales >39, dorsal spines 14-16 .....(17)

17a. Teeth in upper jaw small (more than 50), closely set .....(18)

17b. Teeth in upper jaw large (less than 50), widely spaced .....(19)

18a. Head markedly compressed (narrow), with distinct speckles and wavy bars (vermiculations) ..... *S. angusticeps*

18b. Head moderately compressed, without spots or vermiculations ..... *S. altus*

18c. Head broad, without spots or vermiculations ..... *S. stappersi*

19a. Pectoral fin 19-24 % SL, a dark band along mid body ..... *S. robustus*

19b. Pectoral fin 23.5 – 29% SL, vertical bars on body .....*S. macrocephalus*

19c. Pectoral fin 25.0 – 32% SL, no band or bars on body.....*Serranochromis* **'sp elongate'***

Pharyngochromis species

20a. Black bands on head, SL up to 100mm ..... *P. acuticeps*

20b. No black bands on head, SL >100mm ..... *P.* **'sp elongatus'* (Zambezi drainage)**, *P.***'sp elongatus chambeshi'* (Luapula drainage)**

* Undescribed (new) species

*The undescribed *Sargochromis* species have not been fitted to this key because there is need to use more specimens, than available during this study, to establish their distinctness to a reasonable level beforehand.

4.6. Origin and speciation of the riverine haplochromine cichlids

Haplochromine cichlids form the most species-rich lineage of cichlid fishes that both colonized almost all river systems in Africa and radiated into species flocks in several East and South Eastern African lakes. The enormous diversity of lakes is contrasted by relatively poor species diversity in rivers. The species of the genus *Pseudocrenilabrus* and the serranochromine species belong to the haplochromine cichlids. This group belongs to the H-lineagae of the Lake Tanganyika cichlid radiation, which originated during the "primary radiation" of the species flock (Salzburger *et al.*, 2002). The haplochromine cichlid group comprises both the species of Lake Tanganyika from the tribe Haplochromini (Poll, 1986) and several riverine haplochromines distributed all over Africa (Skelton, 1993). The connection between the Zambezi and Congo systems during the late Tertiary has been suggested by Bell-Cross (1965a). The existence of a lake which dried up 20 000 years ago has been suggested as the source of the Zambian riverine haplochromines by Joyce *et al.* (*in press*)

For the *Pseudocrenilabrus* consensus phylogenetic tree (Figure 71) the outgroup taxa used are *P. multicolor* and *P. nicholsi*. These two *Pseudocrenilabrus* species represent the "northern species" that are assumed to be sister to the *P. philander* complex (Greenwood, 1989; Katongo *et al.*, in press). From this tree it is observed that after the outgroup taxa, the most ancestral split is represented by *Pseudocrenilabrus* 'sp haplochromis- orange', a new species from Lake Mweru. The next split is represented by the *P. philander* from Lake Bangweulu and Chambeshi River (Luapula River system). The most recent split is represented by the *P. philander* from the Zambezi and Kafue Rivers and *P. philander* from the Lunzua River. The *P. philander* from the Lunzua comprises two colourmorphs 'blue' and 'yellow' which represent the same mitochondrial haplotype.

For the serranochromine phylogenetic tree (Figure 73) the outgroup taxa used are *Orthochromis polyacanthus* and *Cyclopharynx fwae*. These taxa are assumed to be sister to the serranochromines as seen in the Tanganyika tree (Figures 68 and 68) and from previous studies (Salzburger *et al.*, 2002; Klett and Meyer 2002). From the

serranochromine tree it is observed that after the outgroup taxa, the most ancestral split is represented by *Serranochromis stappersi* from Lake Mweru (Luapula River system) and *S. robustus* from both the Luapula and Zambezi River systems. The next split is represented by a new species, *Serranocromis* ' sp elongate' from Lake Mweru (Luapula River system) and a combination of *S. angusticeps* and *S. altus*. The most recent split is represented by members of the genus *Sargochromis* from both the Zambezi and Luapula River systems, members of the genus *Pharyngochromis* from the Zambezi River system and *Serranochromis macrocephalus* and *S. longimanus* from the Zambezi River system.

The close evolutionary relationship of Central and Eastern African haplochromines with the H-lineage of the Tanganyika flock led Salzburger *et al.* (2002) to suggest that their origin was connected to the radiation of the H-lineage within Lake Tanganyika area, at a time when the lake was not yet fully isolated ecologically from surrounding rivers. It was hypothesized that the Tanganyika radiation swept over to surrounding rivers and also induced faunal changes in there. According to this hypothesis, the most likely route of colonization of Zambian river systems represents the Congo River, so that a northern origin of the Zambian riverine haplochromines seems likely. Both the *Psuedocrenilabrus* and *Serranochromines* phylogenetic trees support this direction of origin. (Figures 71 and 73)

The most ancestral split within the *Psuedocrenilabrus* is represented by the Lake Mweru-Mwatishi River confluence population. The Zambezi-drainage populations appear to be more derived than the Luapula populations, suggesting a bridging process in which a common ancestor of *Pseudocrenilabrus*, originating from the Luapula (Congo system), split into two groups with one of them colonizing the Zambezi and the other one remaining within the Luapula (Figure 71).

Within the Serranochromines, the most ancestral split is represented by *Serranochromis stappersi*, which is only reported from Lake Mweru. Most of the remaining clades comprise species, which are distributed in both the Luapula and Zambezi system, except the genus *Sargochromis* in which *Sargochromis mellandi* has only been reported from the

Luapula with the other *Sargochromis* species only reported from the Zambezi. Similarly, *Serranochromis altus* has only been reported from the Zambezi system (Figure 73). The most likely origin of the Serranochromines is, therefore, the Upper Congo. From there it is likely to have migrated to the Luapula and then bridged into the Zambezi system.

4.7. *Pseudocrenilabrus* taxonomy, phylogeography and speciation

The mitochondrial DNA phylogeny of the *Pseudocrenilabrus* (Figure 71) identifies four major lineages, three of which occur in the Congo system and one in the Zambezi system. Two of the Congo clades (Lake Mweru and Lunzua River) comprise distinct albeit undescribed species while the third Congo clade (Lake Bangweulu and Chambeshi River), together with the Zambezi clade (Zambezi and Kafue Rivers) are morphologically assigned to *Pseudocrenilabrus philander* (Figure 70).

Concerning intra-population variation, most of the populations are morphologically and genetically highly uniform, suggesting severe founder effects and/or bottlenecking during their history. The linearized tree (Figure 53) reflects two diversification events. The first locates at about 3.9% mean sequence divergence and points to an almost simultaneous colonization of the sampled waters. Subsequent regional diversification (with about 1% mean square divergence) occurred contemporaneously within the Kafue and Zambezi Rivers. The clear-cut biogeographic structure points to the dominance of geographic (allopatric) speciation in this lineage of riverine cichlid fishes (Hewitt, 2001).

4.8 Serranochromine cichlids taxonomy, phylogeography and speciation

The results from this study reveal that a number of *Sargochromis* species mainly from the Zambezi system are yet to be described (Figures 35a, 35b, 35c and 73). The genus *Pharyngochromis* which Greenwood, (1992) has suggested to be monospecific, is likely to comprise two or even more species (Figures 45a, 45b, 46, 47, 40, 60 and 61). The present study also identified one yet undescribed species of *Serranochromis* (*Serranochromis* 'sp elongate') from the Luapula (Figures 32a, 32b, 32c, 60 and 61).

The mitochondrial DNA phylogeny of the serranochromine cichlids (the group as defined by Greenwood, (1993)) identifies four major lineages, of which *S. stappersi* occurs only in the Congo system, while the other four (*Serranochromis robustus/S.thumbergi, Serranochromis altus/S. angusticeps, Serranochromis macrocephalus/S.longimanus and Sargochromis/Pharyngochromis*) occur both in the Zambezi and the Congo systems (Figure 73).

In contrast to *Pseudocrenilabrus*, more than one species from each serranochromine genus lives in Zambian rivers. There is, therefore, no clear-cut geographic separation among the serranochromine species suggesting that their speciation may have occurred within river systems, possibly via sympatric or microallopatric speciation. Thus, speciation in this group is likely to have been driven, among other factors, by different dietary preferences influenced by their mouth and jaw types.

Chapter 5 Conclusions and recommendations

5.1 Conclusions

Three major conclusions were drawn from the results of this study:

1.The results from this study suggest that there are likely to be a number of undescribed species and/or misidentifications within the Zambian riverine haplochromine cichlids. For example, there are (most likely) more than one species within the genus *Pseudocrenilabrus* in Zambia. The population from Lake Mweru / Mwatishi River confluence is morphologically the most distant from the other populations.

Moreover, the *Sargochromis* specimens from the Luapula collected by Van Zwieten (MRAC 9-035-P) appear to consist of two species with one of them likely to be *Sargochromis mellandi* while the second is likely to be a different species. The collection of *Sargochromis* from the present study (MRAC A4-25), from Lakes Mweru and Bangweulu also appears to comprise two different species.

There seems to be one or two undescribed species of *Sargochromis* from the Kafue River. One of them, a slender type with a long snout, was quite abundant during the field trip of the summer of 2002, while the second one with longitudinal stripe(s) was less abundant. Van der Waal's collection (RUSI 062863) has one *Sargochromis* species collected from below the Itezhi Tezhi Dam, on the Kafue River. This taxon is very likely to be a new record in the Kafue or even an undescribed species.

The *Pharyngochromis* group seems to consist of more than one species contarary to Greenwood's (1992) assertion. The *Pharyngochromis* from Lake Kariba seems to fit very well into the description of *Pharyngochromis acuticeps* by Skelton (1993) especially in terms of the dark coloration on the head and the generally small size of less than 100 mm. The other *Pharyngochromis* individuals are bigger in size with different coloration on the head. There seem to be some individuals, which are intermediate between *Sargochromis*

and *Pharyngochromis* collected from the Upper Zambezi (Plate 8). Also collected during this study was a *Pharyngochromis*-like specimens from the Chambeshi River, which is certainly different from *Pharyngochromis acuticeps*.

The *Serranochromis* group presents fewer taxonomic problems. However, there is a confusingly close morphological similarity between *Serranochromis macrocephalus, S. stappersii* and a new *Serranochromis* 'sp elongate' from Lake Mweru in the Luapula drainage. This new species has a very close mitochondrial DNA relationship with *S. angusticeps* and *S. altus*. *Serranochromis longimanus* is a relatively rare species and only two specimens were collected during the present study, one of which has been deposited in the Africa Museum (MRAC A4-25).

2. Results obtained from morphological analyses are similar to those obtained from molecular analyses and they seem to complement each other.

3. The riverine haplochromines seem to have colonized the rivers of Zambia from the Upper Congo to the Luapula and finally from the Luapula to the Zambezi drainage.

5.2. Recommendations for future work

Preliminary results from this study indicate that there is a need to revise the genera *Serranochromis* (especially from the Luapula-Congo drainage), *Sargochromis*, *Pseudocrenilabrus* and *Pharyngochromis* from Zambian rivers. The revision of each genus should involve field collections of enough specimens, which should then be subjected to taxonomic and phylogenetic analysis. It would also be interesting to include microsattelite analysis of populations within each species to study the intraspecific variation and geographic separation within each species. *Serranochromis longimanus* and *Chetia* seem to be rare species in Zambian rivers but this should be verified by abundance and distribution studies before the species are listed as rare. Cladistical analysis of riverine haplochromine cichlids based on morphological characters of evolutionary importance would be interesting in this group.

Results from the linearized trees of the *Pseudocrenilabrus* and Serranochromines, respectively should be combined with geological data of the study area to calibrate a molecular clock for the Zambian riverine haplochromines in order to date the major cladogenetic events in the phylogeny of the groups.

References

Adachi, J and M. Hasegawa. 1994. MOPHY. *Programs for molecular phylogenetics.* Version 2.2. Institute of Statistical Mathematics, Tokyo.

Akaike, H.1974. A new look at the statistical model identification. *IEEE Trans. Autom.Contr.* **19**: 716-723.

Alberts, B ; D.Brady ; J.Lewis ; K.Roberts and J.D.Watson.1994.*Molecualr Biology of the Cell.* Garland Publishing, New York.

Ambali, A; H. Kabwazi; H. Malekano; G. Mwale; D. Chimwaza; J. Ingainga; N. Makimoto; S. Nakayama; M. Yuma and Y. Kada. 2001. Relationship between local an scientific names of fishes in Lake Malawi/Nyasa. *African Study Monographs.* **22**(3): 123-154.

Barel, C.D.N; M.J.P. van Oijen; F.Witte and E.L.S. Witte-Maas. 1977. An Introduction to the Taxonomy and Morphology of the Haplochromine Cichlidae from Lake Victoria. (Figures and text). *Netherlands Journal of Zoology* **27(4):** 333-380(1997).

Baric, S.2000. *Impact of Lake Fluctuations on population structure and speciation in Rock dwelling cichlids from Lake Tanganyika.* PhD thesis. University of Innsbruck. Institute of Zoology and Limnology.

Beckenbach AT.1991. Rapid mtDNA sequence analysis of fish populations using the polymerase chain reaction (PCR). *Canadian Journal of Fish and Aquatic Science.* **48** (Supplementary 1).

Beadle, L.C.1981. *The Inland Waters of Tropical Africa-* Longman, London.

Bell-Cross, G. 1965a. The Distribution of Freshwater Fishes of Central Africa. *Fisheries Research Bulletin* .Zambia. **4** : 3-20.

Bell-Cross, G. 1965b. *Movement of fish across the Congo-Zambezi watershed in the Mwinilunga district of Northern Rhodesia.* African Science and Medical Congress, Lusaka. Pergamon, Oxford. Pp 415-424.

Bell-Cross, G. 1966. The Distribution of Fishes in Central Africa. *Fisheries Research Bulletin* **4**: 3-15, Zambia.

Bell-Cross, G. 1972. The Fish Fauna of the Zambezi System.*Arnodia.* (Rhodesia), (**29**) **5** : 1-19.

Bell-Cross, G. 1976. Physical barriers separating the fish fauna of the Kafue and the Middle Zambezi River Systems. *Fisheries Research Bulletin.* **4** : 97-98. Zambia.

Bell-Cross, G. 1975. A revision of certain *Haplochromis* species of Central Africa. *Occasional Papers of the National Museums and Monuments of Rhodesia.*

Birky CW Jr., P. Fuerst and T. Maruyama.1989. Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplastic cells and comparison to nuclear genes. *Genetics* **121**.

Boulenger, G.A. 1915. *Catalogue of Fresh-water Fishes of Africa.* **Volume III.** The British Museum of Natural History. London.

Bowmaker A.P., P.B.N. Jackson and R.A. Jubb. 1978. *Freshwater Fishes. In Biogeography and Ecology of Southern Africa.* (Ed. M.J.A. Werger). Pp. 1181-1230. **31** Junk. The Hague.

Brichard, P. 1989. Cichlids of Lake Tanganyika. T.F.H.Publications. Neptune City.

Brodersen J, MJ Chimbari and H Madsen. *2002*. An enclosure study to evaluate the effect of *Sargochromis codringtonii* on snail populations in an irrigated canal. *African Zoology* **37**.

Brown WM, M JR George and AC Wilson. 1979. Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Science. USA **76**.

Cavalli-Sforsa, L.L. and A.W.F. Edwards. 1967. Phylogenetic analysis: Models and estimation procedures. *Evolution* **35** : 32 : 550-570.

Chimbari MJ, J Ndamba and H Madsen. 1986. Food selection behaviour of potential biological agents to control intermediate hosts of schistomiasis: *Sargochromis codringtonii* and *Tilapia rendalli*. *Acta Tropica* **61**.

Chimbari MJ, H Madsen and J Ndamba. 1987a. Laboratory experiments on snail predation by *Sargochromis codringtonii*, a candidate for control of snails that transmit schistomiasis. *Annals of Tropical Medicine and Parasitology* **91**.

Chimbari MJ, H Madsen and J Ndamba. 1987b. Simulated field trials to evaluate the effect of *Sargochromis codringtonii* and *Tilapia rendalli* on snails in the presence and absence of aquatic plants. *Journal of Applied Ecology* **34**.

Clayton DA. 1982. Replication of animal mitochondrial DNA . *Cell* **28**.

Consuergra, S; C. G. de Leaniz; A. serdio, M. G. Morales; L.G. Straus D. Knox and E. Verspoors. 2002. Mitochondrial DNA variation in leistocene and modern Atlantic salmon from Iberian glacial refugium. *Molecular Ecology*.

Eccles, D.H. and E. Trewavas. 1989. *Malawian cichlid fshes. The classification of some haplochromine genera*. 335pp. Lake Fish Move, Hetren, Germany.

Farias, I.P ; G.Orti and A. Meyer. 2000. Total evidence: Molecules, morphology and the phylogenetics of cichlid fishes. *Molecular Development and Evolution*. **288**: 76-92.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* **17**.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783-791.

Felsenstein, J. 1988. Phylogenies from molecular sequences: inference and reliability. *Annual Review in Genetics* **22**.

Felsenstein, J. 1993. *PHYLIP: Phylogenetic inference package*. Version 3.5c. Department of Genetics, University of Washington, Seattle.

Felsenstein, J. and H. Kishino.1993. Is there something wrong with bootstrap on phylogenies? A reply to Hillis and Bull .*Bullettin of Systematics Biology* **42**: 193-200.

Fryer, J. and T.D. Iles. 1972. *The cichlid fishes of the Great Lakes of Africa – their biology and evolution*. Oliver and Boyd, Edinburgh.

Froese, R. and D. Pauly, 2000. *Fishbase 2000*. ICLARM, Phillipines and Fisheries Center, UBC, Vancouver, Canada.

Gaigher, I.G and Pott, R, Mc.1973. Distribution of fishes in Southern Africa. *South African Journal of Science*. **60:** 383-391.

Gassen, H.G; G.E.Sachse and A.Schulte. 1994. *PCR: grundlagen und Anwendugen der Polymerase-Kettenreaktion.* Gustav Fisher Verlag. Stuttgart, Jena, New York.

Goldstein, D.B and D.D. Pollock. 1994. Least squares estimations of molecular distance noise abatement in phylogenetic reconstruction. *Theoretical Population Biology.* **45:** 219-226.

Greenwood, P.H. 1978. A review of the pharyngeal apophysis and its significance in the classification of African cichlid fishes. *Bulletin of the British Museum of Natural History (Zoology).* **33:** 297-323.

Greenwood, P.H. 1984. African cichlid fishes and evolutionary theories. Pp 121-154 in A.A. Echelle and I. Kornfield, Eds. *Evolution of fish species flocks.* Univ. of Main at Orono Press, Orono.

Greenwood, P.H. 1989. The taxonomic status and phylogenetic relationships of *Pseudocrenilabrus* Fowler (Tolestei, Cichlidae). *Ichthyological Bulletin of the J.L.B. Smith Institute of Ichthyology,* Grahamstown No. **54**:1-16.

Greenwood, P.H.1993. A review of the Serranochromine genera, *Pharyngochromis,* Sargochromis, Serranochromis and Chetia. Bulletin of the Natural History Museum, **59.**

Hasegawa, M; H. Kishino and K. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution.* **22.**

Hasegawa, M and M. Fujiwara. 1993. Relative efficiencies of maximum likelihhod, maximum parsimony and distance methods for estimating protein phylogeny. *Molecular Phylogenenetics and Evolution.* **2** : 1-5.

Hartl, D.L.2000. *A primer of population genetics.* Third edition. Sinauer Associates, Sunderland, Massachusetts.

Hedges, S.B. 1992. The number of replications needed for accurate estimation of the bootstrap P value in phylogenetic studies. *Molecular Biology and Evolution.* **9** : 366-369.

Hennig, W. 1965. Phylogenetic Systematics. Annu. Rev. Entomol. **10**: 97-116.

Hennig, W. 1966. Phylogenetic Systematics. University of Illinois Press. Urbana.

Hewitt, G.M. 2001. Speciation, hybrid zones and phylogeography-or seeing genes in space and time. Molecular Ecology.

Higashi, M ; G. Takimoto and N. Yamamura. 1999. Sympatric speciation by sexual selection. *Nature*. Macmillan Magazines Ltd.

Hillis, D.M. and J.J. Bull.1993. An empirical test of bootstrapping as a method of assessing confidence in phylogenetic analysis analysis. *Systematics Biology*. **42:** 182-192.

Hills, D.M.; C. Moritz and B.K. Mable. 1996 (Eds). Chapter 11. *Molecular Systematics*. 2nd Edition. Sinauer Associates, Inc. Sunderland, Massachusetts USA.

Huelsenbeck, J.P. 1995. The robustness of two phylogenetic methods: Four-taxon simulations reveal a slight superiority of maximum likelihood over neighbour joining. *Molecular Biology and Evolution.* **12**: 843-849.

Jackson, P.B.N. 1961. The Fishes of Northern Rhodesia. A Check List of Indigenuos Species. Government Printer, Lusaka, Zambia.

oyce, D.A; D.H. Lunt; R. Bills; G.F. Turner; C. Katongo; N. Duftner; Christian turmbauer and O. Seehausen. (*In press*). An extant cichlid fish radiation emerged in an xtinct Pleistocene lake. *Nature*.

ıbb, R.A. 1964. Freshwater fishes and drainage basins in Southern Africa. *South African ournal of Science*. **60**: 17-21.

bb, R.A. 1967. *Freshwater fishes of Southern Africa*. Balkema, CapeTown.

Kai, J; K. Nakayama and T. Nakabo. 2002. Genetic differences among three colour morphotypes of the black rockish, *Sebastes inermis*, inferred from mtDNA and AFLP analyses. *Molecular Ecology.*

Katongo,C; S. Koblmueller; N.Duftner; L. Makasa and C.Sturmbauer.(*In press*). Phylogeography and speciation in the *Pseudocrenilabrus philander* species complex in Zambian rivers. *Hydrobiologia.*

Kenmuir, D.1983. *Fishes of Kariba.* Longman, Zimbabwe.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences *Journal of Molecular Evolution* **16**.11 1-20.

Kishino, H., T. Miyata and M. Hasegawa. 1990. Maximum likelihood inference of of protein phylogeny and origin of chloroplasts. *Journal of Molecular Evolution.* **31**: 151-160.

Kocher T.D; Thomas W.K; A. Meyer; S.V. Edwards; S.Paabo ; F.X. Villablanca and A.C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals : amplification and sequencing with conserved primers. *Proceedings of the National Academy of Science.* USA **86.**

Kocher T.D; J.A Conroy; K.R McKaye and J.R Stauffer.1993. Similar morphologies of Cichlid fish in Lakes Tanganyika and Malawi are due to convergence. *Molecular Phylogenetics and Evolution* **2.**

Kohda, M; Y. Yanagisawa; T. Sata; K. Nakaya; Y. Niimura, K. Matsumoto and H. Ochi. 1996. Geographic colour variation in cichlid fishes at the southern end of Lake Tanganyika. *Environmental Biology of Fishes.*