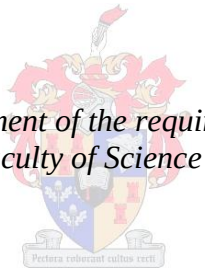
A molecular phylogenetic study of the South American fish genus *Pterophyllum*

by
Chanelle Anderson



Thesis presented in fulfilment of the requirements for the degree of Master of Science in the Faculty of Science at Stellenbosch University

Supervisor: Prof. Dirk Uwe Bellstedt

March 2016

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Chanelle Anderson

March 2016

ABSTRACT

The South American fish genus *Pterophyllum*, commonly referred to as Angel Fish, contains three recognized species namely *Pterophyllum leopoldi*, *Pterophyllum altum* and *Pterophyllum scalare*. In addition, morphological variants, particularly of *Pt. scalare* and *Pt. altum*, occur in their natural habitat. However, uncertainty regarding the taxonomy and the phylogenetic relationships between the three species of the genus *Pterophyllum* within the tribe Heroini, still remains.

In the first part of this study three mitochondrial (Cytochrome *b*, 16S rRNA and NADH dehydrogenase subunit 4) and two nuclear (Recombinase Activator Gene 2 and ribosomal S7 intron 1) markers were amplified and sequenced for 27 *Pterophyllum* samples. Along with this, 166 nucleotide reference sequences, representing six outgroups and 160 Neotropical cichlid taxa, were downloaded for each of the five molecular markers from GenBank. In addition to maximum likelihood and parsimony analyses, a BEAST analysis was performed on the Cytochrome *b* sequence matrix. This study was the first to conclusively show that the genus *Pterophyllum* is a monophyletic group of which *Pt. leopoldi* is the basal lineage whilst *Pt. altum* and *Pt. scalare* form a derived sister association. Furthermore, it appears that the evolution of the genus *Pterophyllum* corresponds to the historical geographical changes of the Amazonian landscape. In addition, the study also provides phylogenetic evidence for the reconnection of the Amazon and Orinoco river systems by means of the Casiquiare channel.

In the second part of this study the sequence variability of 13 additional nuclear markers were investigated in order to expand the five-gene analysis with the aim of enhancing the resolution of the phylogeny. In conjunction with this, the number of *Pterophyllum* taxa included in the ND4 sequence matrix were increased from two to six representatives for each population of the three species in order to elucidate intraspecies and population variability. The two nuclear genes Mitfb and RpL8 were identified as the most variable of the 13 nuclear genes investigated. The inclusion of these two nuclear genes in the five-gene data matrix, improved the ratio of mitochondrial to nuclear variable characters from 79:21 to 67:33 in the seven-gene phylogeny. Distinction between *Pt. altum* of the upper Rio Negro and Upper Rio Orinoco was repeatedly confirmed with full statistical support. Variation within *Pt. scalare* retrieved three groups of which *Pt. scalare* Santa Isabel and *Pt. scalare* Manacapuru were the basal lineages. This study offers no phylogenetic evidence to support a hybrid origin for the latter.

OPSOMMING

Die Suid-Amerikaanse vis genus *Pterophyllum*, wat algemeen verwys word na as “Angelfish”, bevat tans drie herkende spesies naamlik *Pterophyllum leopoldi*, *Pterophyllum altum* en *Pterophyllum scalare*. Tans heers daar egter onsekerheid oor die taksonomiese posisie sowel as die filogenetiese verwantskappe tussen die drie spesies van die genus *Pterophyllum*, binne die tribus Heroini, aangesien morfologiese variante, veral van *Pt. scalare* en *Pt. altum*, in hul natuurlike habitat voorkom.

In die eerste deel van hierdie studie was drie mitokondriale (Sitochroom *b*, 16S rRNS en NADH dehidrogenase subeenheid 4) en twee nuklêre (Rekombinasie Aktiveringsgeen 2 en ribosomale S7 intron 1) merkers vir 27 *Pterophyllum* monsters ge-amplifiseer en die nukleotied volgordes daarvan bepaal. Tesame hiermee was 166 nukleotied verwysingsvolgordes verteenwoordigend van ses buitegroep en 160 Neotropiese “cichlid” spesies vir elk van die vyf molekulêre merkers van GenBank afgelaai. Addisioneel tot “maximum likelihood” en “parsimony” analises, was ’n BEAST analise op slegs die Sitochroom *b* volgorde matriks uitgevoer. Hierdie studie was die eerste om onweerlegbaar te toon dat die genus *Pterophyllum* ’n monofiletiese groep vorm waarvan *Pt. leopoldi* basaal en *Pt. altum* en *Pt. scalare* in ’n afgeleide sister verhouding tot mekaar staan. Verder bleik die evolusie van die genus *Pterophyllum* ooreen te stem met die geskiedkundige geografiese veranderinge van die Amasone-streek. Daarmee saam bied die studie ook filogenetiese bewyse vir die heraansluiting van die Amasone en Orinoco rivier sisteme deur middel van die Casiquiare kanaal.

In die tweede deel van hierdie studie was 13 addisionele nuklêre merkers se volgorde variabiliteit ondersoek ten doel die vyf-geen analise aan te vul om sodoende die resolusie van die filogenie te versterk. Terselfdertyd was die *Pterophyllum* taksa in die ND4 volgorde matriks uitgeberei van twee tot ses verteenwoordigers per populasie van elke spesie om sodoende intraspesies en populasie variabiliteit te ondersoek. Die twee nuklêre gene, Mitfb en RpL8, was as die mees variabele nuklêre gene geïdentifiseer. Die insluiting van hierdie twee gene by die vyf-gene data matriks het die verhouding mitochondriale tot nuklêre variabele karakters van 79:21 na 67:33 in die sewe-geen filogenie verbeter. Onderskeid tussen *Pt. altum* van die bo Rio Negro en bo Rio Orinoco kon herhaaldelik met volle statistiese steun bevestig word. Variasie binne *Pt. scalare* het drie groeperinge getoon waarvan die *Pt. scalare* Santa Isabel en *Pt. scalare* Manacapuru basale klades was. Die studie bied egter geen filogenetiese bewyse in steun van ’n hibried oorsprong vir laasgenoemdes.

ACKNOWLEDGEMENTS

I would like to sincerely thank the following people and institutions for their contribution to this study:

My supervisor, **Professor Dirk U Bellstedt**, for making this project available and your guidance throughout. It has been a privilege to learn from you.

Dr Woody Cotterill for your assistance with the BEAST analysis.

Mr De Wet van der Merwe for your assistance with the phylogenetic analyses and guidance in general.

Mrs Coral de Villiers for your great management of the laboratory.

Mrs Welma Maart for your technical support.

The individuals of both the **Bellstedt and Botes laboratories** for your advice, support and encouragement.

My parents, **Danie** and **Ester Anderson**, for your love and support.

My sister **Anecia Conradie**; a special thanks to you for ALL your help, love, continued support and patience.

My dear friends **Vicus Kruger**, **Bertie de Wet**, **Lindie Smit** and **Christilee Jansen**.

The **National Research Foundation** (NRF) for financial support.

ABBREVIATIONS

AIC	Akaike Information Criterion
AICc	Akaike Information Criterion (corrected)
BEAST	Bayesian Evolusionary Analysis by Sampling Trees
BEAUti	Bayesian Evolusionary Analysis Utility
BIC	Bayesian Information Criterion
CI	Consistency Index
DT	Decision Theory criterion
EDTA	Ethylenediaminetetraacetic acid
ESS	estimated sample size
GTR	General Time Reversible
HPD	highest posterior denisty
iTOL	interactive tree of life
MCMC	Markov Chain Monte Carlo
ML	maximum likelihood
MP	parsimony
MRCA	most recent common ancestor
Ma	Mega-annum (million years)
NaCl	Sodium Chloride
PAUP*	Phylogenetic Analysis Using Parsimony
PCI	Phenol-Chloroform-Isoamylalcohol
PCR	Polymerase chain reaction
Pt.	*Pterophyllum*
RI	retention index
RO	reverse osmosis
SDS	sodium dodecyl sulfate
ST_eP	stepped elongation time protocol
TAE buffer	Tris-acetate-EDTA buffer
TBR	Tree Bisection and Reconnection
TE-buffer	Tris-EDTA buffer
tmrca	time to most recent common ancestor
Tris	tris(hydroxymethyl)aminomethane
XML	Extensible Markup Language
XSEDE	Extreme Science and Engineering Discovery Environment

TABLE OF CONTENTS

1. INTRODUCTION

The South American genus *Pterophyllum*, commonly known as Angelfish, consists of three valid species; *Pterophyllum scalare* (Schultze, 1823), *Pterophyllum altum* (Pellegrin, 1903) and *Pterophyllum leopoldi* (Gosse, 1963). Their unique morphological features include being shortened and laterally compressed. However, it is their extremely elongated dorsal and ventral fins that are their most characteristic physical feature and which distinguish them from any other cichlid fish. Studies by Stiassny (1991), Kullander (1998), Farias et al. (2000), Smith et al. (2008) and López-Fernández et al. (2010) have contributed significantly to understanding the phylogenetic relationships and thus the evolutionary history of the Neotropical cichlids. However, despite some degree of congruency among their hypothesised topologies for the Neotropical subfamily, the taxonomic position of the genus *Pterophyllum* in relation to its related genera in the tribe Heroini remains uncertain. Thus, inspite of the comprehensive knowledge of the genus *Pterophyllum* and the fact that it contains three recognised species, it has been underrepresented at species-level.

Pterophyllum species are one of the most popular ornamental fish besides members of the genus *Symphysodon* commonly referred to as discus. For this reason *Pterophyllum*, in particular *Pt. altum*, is of great interest in aquaculture and the ornamental fish trade where they are widely exploited (Schneider et al., 2012). Fishkeeping is a popular hobby worldwide and therefore a well-established and lucrative industry with significant annual exports of Neotropical cichlid fish from the Amazon and Orinoco rivers. Biogeographically, the genus *Pterophyllum* is widespread throughout the Amazon River, the Orinoco River and the Essequibo River drainage basins. Of the three species, *Pt. altum* is the most challenging to maintain due to its very specific requirements such as very soft water with a low pH; conditions which conform to those in its natural environment. *Pterophyllum altum* species are therefore of high economic value and a high commercial demand exists for this species. Recently, a few breeders, particularly in Germany, have had success in breeding *Pt. altum* (Forkel, 2015). In contrast *Pt. scalare* is the most common representative of the genus *Pterophyllum* and can be found in almost every pet shop and breeder's aquaria. Breeders have successfully bred with these species without much difficulty; hence they have become the 'common' species. Last, but not the least, *Pt. leopoldi* appears to be the species that draws the least attention. Very few, if any, cases involving the keeping and breeding of this species have been reported. Recently, some intermediate morphological variants of *Pterophyllum* have been discovered and there is uncertainty about which *Pterophyllum* species they can be assigned to.

Consequently, the objectives of this study were to:

- Establish the phylogenetic relationships between the three *Pterophyllum* species and test whether the genus *Pterophyllum* is monophyletic.

- Determine the taxonomic position of the genus *Pterophyllum* within the tribe Heroini and the Neotropical Cichlidae.
- Investigate whether the evolution of the three species of the genus *Pterophyllum* can be explained based on changes in the Amazonian landscape and its drainage patterns through time.
- Investigate whether the fish collections from the upper Rio Negro in Santa Isabel and Lago Manacapuru (within the vicinity of Manaus) which have an intermediate morphology between *Pt. scalare* and *Pt. altum* are of hybrid origin.

The research presented in this thesis is thus focussed on the Neotropical freshwater genus *Pterophyllum* and aims to contribute to the improved understanding of its evolutionary history and the phylogenetic relationships between its three valid species *Pt. altum*, *Pt. scalare* and *Pt. leopoldi*. In Chapter 2 a literature review of the phylogenetic history of the family Cichlidae, Cichlinae and tribe Heroini is presented. This is followed by an indepth review of the morphological differences between the three species of the genus *Pterophyllum* as well as their phylogenetic history. Furthermore, in light of the biogeographic distribution of the three species of the genus *Pterophyllum*, the changes in the Amazonian landscape and drainage patterns through time was reviewed. Finally, certain aspects of molecular phylogenetics, as it applies to this study, is discussed. In Chapter 3, the nucleotide sequences of 27 *Pterophyllum* taxa, representative of five phylogenetic markers (Cyt *b*, 16S rRNA, ND4, RAG2 and S7) were determined and, in combination with those of the 166 Neotropical reference taxa, used to generate a five-gene phylogeny for the Neotropical Cichlinae. In addition, a time-calibrated Bayesian phylogeny, based solely on the mitochondrial Cyt *b* gene, was generated for the Neotropical Cichlinae. In Chapter 4, the sequence variability of thirteen additional and possibly variable nuclear markers was investigated for the purpose of identifying more variable nuclear markers. In addition, the sequence matrix of the most variable phylogenetic marker, ND4, was expanded by increasing the number of representatives of the three species of the genus *Pterophyllum* to investigate intraspecies variability. The conclusions and future perspectives of this study are presented in Chapter 5 after which references and relevant appendices follow.

2. LITERATURE OVERVIEW

2.1 FAMILY: CICHLIDAE

The freshwater fish family Cichlidae (order: Perciformes, suborder: Labroidei) is one of the most species-rich fish families. This family was first described by Regan (1905, 1906). Following his cursory revision of the family Cichlidae, many researchers have attempted to resolve the higher-level intrafamilial relationships of the family. To this end, the work by Kullander (1983, 1986, 1998) and Stiassny (1987, 1991) have arguably, contributed most significantly. Following his influential work in 1983, Kullander (1986) published the book entitled "*Cichlid fishes of the Amazon River drainage of Peru*" in which he provided a "*provisional list of South American genera*." This list contained detailed descriptions of 275 species of which at least 70 species were described for the first time (Kullander, 1986). In turn Stiassny focussed on the resolution of the phylogenetic relationships and subsequent taxonomic placement of the genus *Cichla* (Stiassny, 1987). In her review of cichlid intrafamilial relationships (Stiassny, 1991), Stiassny provided a summary cladogram (her figure 1.3) depicting the phylogenetic relationships among the major cichlid lineages (**Figure 2.1**). Almost a decade later, Kullander proposed the first acknowledged, fully resolved phylogeny and classification of the family Cichlidae entitled: "*A phylogeny and classification of the South American Cichlidae*" (Kullander, 1998).

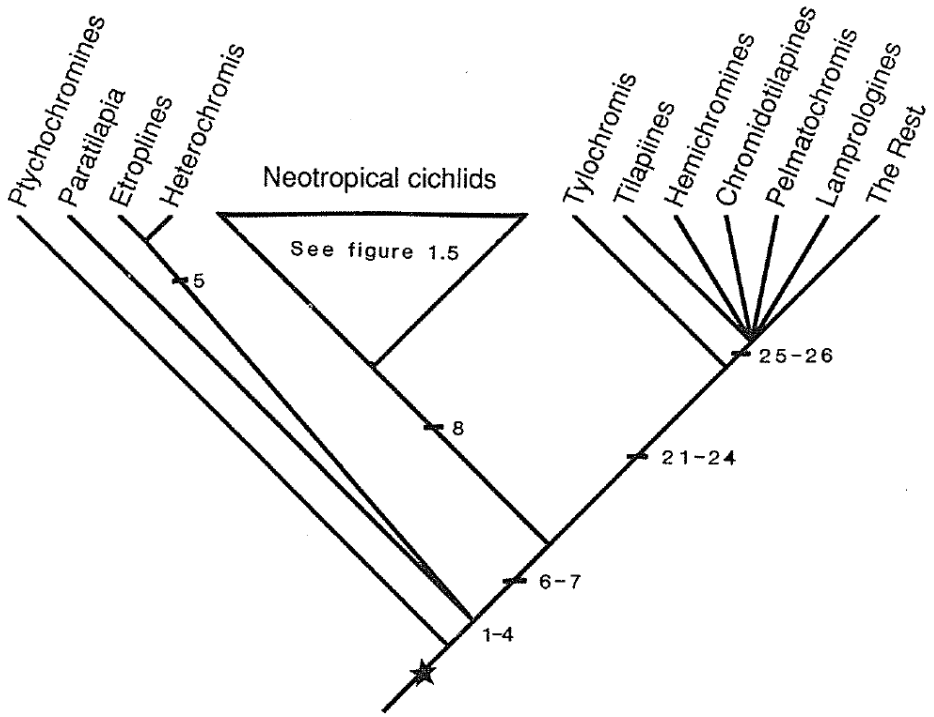


FIGURE 2.1 Summary cladogram of cichlid intrafamilial relationships as proposed by Stiassny, (1991)**.** The numeric values at the base of some clades correspond to the 28 morphological characters used to group the respective cichlid genera. Descriptions of these morphological characters were discussed by Stiassny under the section "taxonomic appendix" (Stiassny, 1991).

However, attempts to resolve the relationships of the family Cichlidae up to the early 1990's, including the work by Kullander (1998), were solely based on morphological characteristics. These studies included the elements of the pharyngeal jaw apparatus which allows for the efficient crushing and processing of prey. According to Karl & Streelman (1997) the use of these elements was unreliable in the construction of evolutionary hypotheses and therefore unreliable as markers of perciform evolution. This was reiterated by the combined results of the three independent studies of Concheiro Pérez et al. (2007), Rícan et al. (2008) and López-Fernández et al. (2010). These studies were in agreement that the use of traditional morphological characters "*to define either the Central American cichlid genera or sections of the former Cichlasoma, as described by Regan 1905, are generally subject to pervasive homoplasy and are thus misleading when used as the sole information to define evolutionary lineages*" (López-Fernández et al., 2010). Nevertheless, the work by Stiassny and Kullander, amongs others, initiated the unraveling of, and laid the platform for the subsequent refinement of the very complex phylogeny of the family Cichlidae. In doing so, they have made a significant contribution to the fields of piscine systematics, taxonomy and phylogeny.

Up to the middle to late 1990's there was still uncertainty about the phylogenetic relationships between the four major cichlid lineages. During this time, researchers made use of the microsatellite flanking region, Tmo-M27 and single-copy nuclear DNA, Tmo 4C4 in order to infer phylogenetic hypotheses for the family Cichlidae (Farias et al., 1999; Karl and Streelman, 1997; Streelman et al., 1998; Zardoya et al., 1996a). Studies by Zardoya et al. (1996), Karl & Streelman (1997) and Streelman et al. (1998) placed particular emphasis on the African cichlids. However, they failed to include the African genus *Heterochromis* and lacked comprehensive taxon sampling of the Neotropical lineage. In contrast, the seminal work by Farias et al. (1999) has made one of the most significant breakthroughs in cichlid phylogeny as they recovered the controversial genus *Heterochromis* at the base of a monophyletic African clade (**Figure 2.2**).

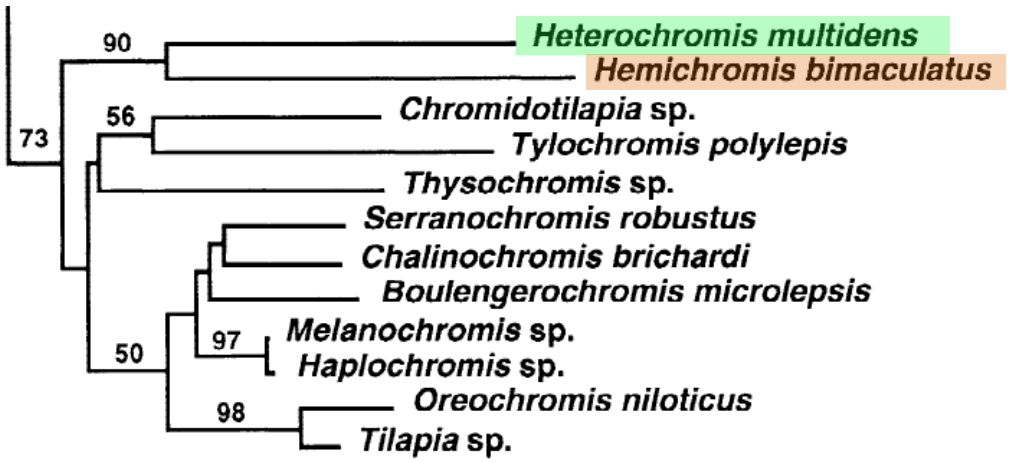


FIGURE 2.2 The monophyletic African clade as recovered by Farias et al., (1999). The clade consisted of 12 species of which the closely related genera *Heterochromis* (highlighted in green) and *Hemichromis* (highlighted in orange) were recovered as the basal lineage of the African clade. Values indicated above branches correspond to bootstrap values.

During the 20th century researchers truly began to formulate robust phylogenetic hypotheses for the family Cichlidae as their studies were based on both molecular and total evidence analyses. The latter involved the use of both morphological and molecular data sets to infer phylogenetic hypotheses which are known to produce better resolved and supported topologies (Lopez-Fernandez et al., 2005). Accordingly, due to the solid foundation laid by the above mentioned researchers in combination with advancements in technology, that is to say improved analytical methods and computational tools, the intrafamilial relationships in the family Cichlidae have been determined with much greater accuracy.

The Cichlidae is the most species-rich family of freshwater fishes (Zardoya et al. 1996a; Kullander 2003) with more than 1600 valid species (McMahan et al., 2013). The vast majority of the family's species are freshwater fishes with a few exceptions that can tolerate brackish and even salt water (Murray, 2000). It is estimated that the number of species is approximately 2000 (Brawand et al., 2014; Farias et al., 2000; Kullander, 2003), as it is likely that there are still many undescribed cichlid species and hence this increased estimate by comparison with the former estimate of 1600 species (McMahan et al., 2013). Although still controversial, an early Cretaceous origin is hypothesised for the family dating back at least 130 million years (Ma) ago, prior to the break-up of the supercontinent Gondwana.

The Cichlidae comprises four sub-families: the Pseudocrenilabrinae, which is distributed across Africa, the Middle East and Iran; the Cichlinae, which inhabits the Neotropics from the South of Texas across Mesoamerica, Central America, Cuba, Hispaniola and South America; the Etroplinae, which can be found in Southern India and Sri-Lanka and finally the Ptychochrominae, which is endemic to Madagascar (Kullander 2003; Sparks & Smith 2004; McMahan et al. 2013). Of the four subfamilies, the Pseudocrenilabrinae is the largest with more than 1000 valid species (McMahan et al., 2013). The latter is also referred to as the Old World Cichlids as they inhabit the freshwaters of the East African Lakes, collectively known as the African Rift Valley lakes which include Lake Malawi, Lake Tanganyika and Lake Victoria. Together, the African Rift Valley lakes form the center of Afrotropical cichlid biodiversity (Farias et al., 2000; Zardoya et al., 1996a). The Neotropical subfamily Cichlinae, which is commonly known as the New World Cichlids, is the second largest with just over 500 valid species (McMahan et al., 2013). The remaining two subfamilies, Etroplinae and Ptychochrominae, are in the minority with only 16 and 15 valid species respectively (McMahan et al., 2013).

For the past two decades researchers have been in agreement about the phylogenetic relationships between the four major lineages of the family Cichlidae. Their conclusions are consistent in the placement of the Malagasy/Indian cichlids as the basal lineage of the family, whilst the African and Neotropical cichlids are monophyletic and sister to one another (Farias et al., 2000, 1999; Karl and Streelman, 1997; Sparks, 2004; Streelman et al., 1998; Zardoya et al., 1996a). However, relationships at the genus level remain unresolved and disputed, particularly within the Neotropical clade (Kullander, 2003). Due to their significant behavioural, ecological and morphological diversity,

cichlids have been the focus of many evolutionary studies and have subsequently become model organisms for understanding the dynamics of adaptive radiation and speciation among vertebrates (Burress, 2015; Farias et al., 2000, 1999; Karl and Streelman, 1997; Sparks, 2004; Streelman et al., 1998; Zardoya et al., 1996a).

2.2 SUBFAMILY CICHLINAE (NEOTROPICAL CICHLIDS)

The Neotropical cichlids are geographically widespread and are well represented in Central and South America comprising approximately 60 genera with an estimate of at least 600 species (López-Fernández et al., 2010) of which 526 are valid (McMahan et al., 2013). Accordingly, the subfamily is considered the third most diverse freshwater fish fauna family within the Neotropics (López-Fernández et al., 2010). The Neotropical assemblage constitute seven tribes; Retroculus, Cichlini, Astronotini, Chaetobranchini, Geophagini, Cichlasomatini and Heroini (López-Fernández et al., 2010; McMahan et al., 2013; Smith et al., 2008). Of these, the four tribes Geophagini, Heroini, Cichlasomatini and Chaetobranchini form the two major Neotropical lineages: Chaetobranchini-Geophagini and Cichlasomatini-Heroini (Smith et al., 2008). The tribe Geophagini is the oldest, most species-rich and ecomorphologically diverse tribe and its total number of species is estimated to exceed 300 (López-Fernández et al., 2013) of which 238 are valid (McMahan et al., 2013). The tribe Heroini is the second largest with approximately 150 species followed by the tribe Cichlasomatini with 115 valid species (López-Fernández et al., 2013; McMahan et al., 2013).

Neotropical cichlids are dispersed as far north as the lower Rio Grande in Texas, USA, to where they are predominantly present throughout the major river drainages of South America. The latter include, but are not limited to, the Amazon, Orinoco and Essequibo Rivers. Despite having substantially fewer representatives than their well-studied lacustrine sisters, the Neotropical cichlids appear to be genetically, morphologically and ecologically more diverse (Farias et al., 1999). Hence they are known as the World's richest freshwater biota (Musilová et al., 2008) who's evolution has been influenced by many different biotic and abiotic factors. According to Farias et al. (2000), their diversity can be attributed to a lower rate of extinction and on-going speciation which allowed for the conservation of primitive characteristics leading to their increased genetic diversity. However, it is because of this diversity that the phylogenetic relationships between the Neotropical cichlids have not yet been resolved which constrains research into the family's taxonomy, ecological and evolutionary history (López-Fernández et al., 2010).

2.2.1 EVOLUTIONARY HISTORY OF THE NEOTROPICAL ASSEMBLAGE

2.2.1.1 Stiassny

In her 1991 publication entitled "*Phylogenetic intrarelationships of the family Cichlidae: An overview*", Stiassny proposed three alternative hypotheses for the interrelationships in the monophyletic Neotropical lineage (**Figure 2.3**). The Neotropical clade consisted of seven subgroupings: two genera (*Astronotus* and *Cichla*) and five suprageneric groups (Chaetobranchines, Crenicichlines, Geophagines, Cichlasomines Group A and Cichlasomines Group B) (Stiassny, 1991). The Cichlasomines Group A consisted of 11 genera (*Acaronia*, *Caqueteia*, *Hoplarchus*, *Heros*, *Herotilapia*, *Hypselecara*, *Mesonauta*, *Neetroplus*, *Pterophyllum*, *Symphysodon* and *Uaru*) including the informal assemblage 'Cichlasoma'. The latter was grouped into eight sections: '*C. amphilophus*', '*C. archocentrus*', '*C. insertae sedis*', '*C. herichthys*', '*C. nandopsis*', '*C. paraneetroplus*', '*C. theraps*' and '*C. thorichthys*' (Stiassny, 1991). In contrast, the Cichlasomines Group B was comprised of the seven genera '*Aequidens*', *Bujurquina*, *Laetacara*, *Nannacara*, *Tahuantinsuyoa*, *Aequidens sensu stricto* and *Cichlasoma sensu stricto*.

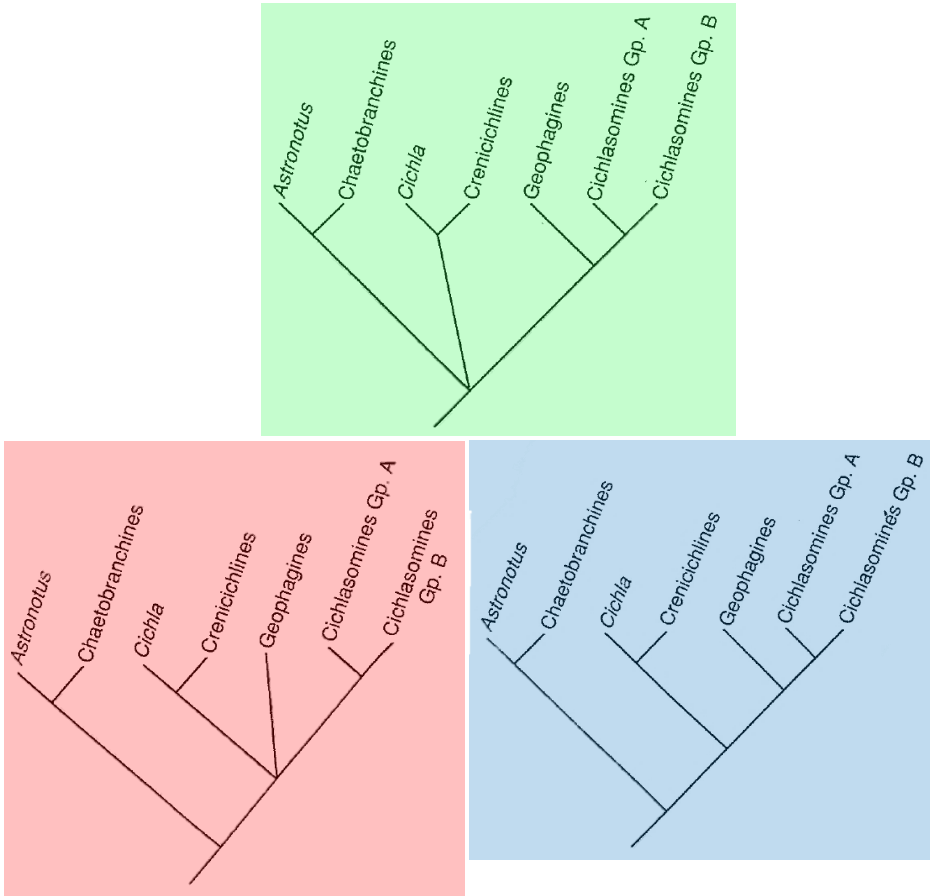


FIGURE 2.3 Three alternative schemes representative of the interrelationships of the Neotropical lineage as proposed by Stiassny (1991)**.** This result was obtained following Farris' successive weighting procedure. In each of the three alternative trees, sister associations between the genus *Astronotus* and the suprageneric group Chaetobranchines in addition to the Cichlasomines Group A and B were retrieved. However, the trees differed in their resolution of the interrelationships of the seven subgroupings.

However, uncertainty about the phylogenetic relationships between the major Cichlidae lineages remained and the family's ecological and morphological characteristics had to be investigated further (Stiassny, 1991). The Neotropical assemblage was thus in need of a robust phylogenetic hypothesis and stable classification for the group. At the time, the areas of particular interest were the unresolved tribe Cichlasomatini and the phylogenetic positions of the genera *Cichlasoma* and *Cichla*.

2.2.1.2 Kullander

In his 1998 publication: "*A Phylogeny and classification of the South American Cichlidae"*, Kullander formulated a phylogenetic hypothesis for the Neotropical assemblage based on the morphological analysis of 91 characters in context of 50 terminal taxa (Kullander, 1998). In his classification, the Neotropical assemblage was subdivided into six subfamilies: Retroculinae, Cichlinae, Heterochromidinae, Astronotinae, Geophaginae and Cichlasomatinae (**Figure 2.4**). However, the inclusion of the controversial African genus *Heterochromis* rendered the Neotropical assemblage paraphyletic. Two major subfamilies were identified: Geophaginae and Cichlasomatinae; each constituting three tribes. Geophaginae included 16 genera grouped within the three tribes Acarichthyini, Crenicaratini and Geophagini. Cichlasomatinae included more than 25 genera grouped within the three tribes Acaroniini, Heroini and Cichlasomatini. Furthermore, the genus *Retroculus* was recovered as the most basal taxon of the Neotropical assemblage followed by the close phylogenetic relationship between the two genera *Cichla* and *Crenicichla*. Also of interest was the close phylogenetic relationship between the two genera *Cichlasoma* and *Aequidens* within the tribe Cichlasomatini. Although well resolved, the exclusion of both Central American and Greater Antillean Heroine cichlids therefore left an inaccurate representation of the evolutionary relationships within the tribe. Probably the most surprising result was the inclusion of the controversial African genus *Heterochromis* within the Neotropical assemblage.

2.2.1.3 Farias et al.

Almost two years after the work by Kullander (1998), Farias et al. (2000) inferred a robust phylogenetic hypothesis for the family Cichlidae based on a total evidence analysis with emphasis on the taxon sampling of Neotropical cichlids (**Figure 2.5**). The study was the most comprehensive at the time as it consisted of four independent analyses; mitochondrial (including 74 taxa), nuclear (including 50 taxa), combined molecular (including 48) and total evidence (including 34 taxa). The total evidence analysis was based on the data sets of one mitochondrial (16S rRNA) and two nuclear (Tmo-M27 and Tmo-4C4) markers in combination with the morphological data set as implemented and published by Kullander (1998). In comparison with the strict consensus tree constructed by Kullander (1998), a few key differences can be highlighted.

First and foremost was the phylogenetic position of the African genus *Heterochromis* at the base of the African radiation. However, this was merely a confirmation of what had been concluded in their

previous study (Farias et al., 1999). With regards to the Neotropical cichlids, the three genera *Retroculus*, *Cichla* and *Astronotus* were recovered as the basal lineages of the assemblage. Furthermore, the genus *Crenicichla* was in close phylogenetic relationship with *Apistogramma* and positioned at the base of the suprageneric group Geophagines. The latter corresponds to Kullander's classification of the subfamily he referred to as Geophaginae. Finally, the genus *Acaronia* was grouped among the Cichlasomines (Kullander's Cichlasomatini) and *Chaetobranchus* was elevated to the suprageneric group Chaetobranchines. In turn the Chaetobranchines were recovered as the sister group to the Geophagines.

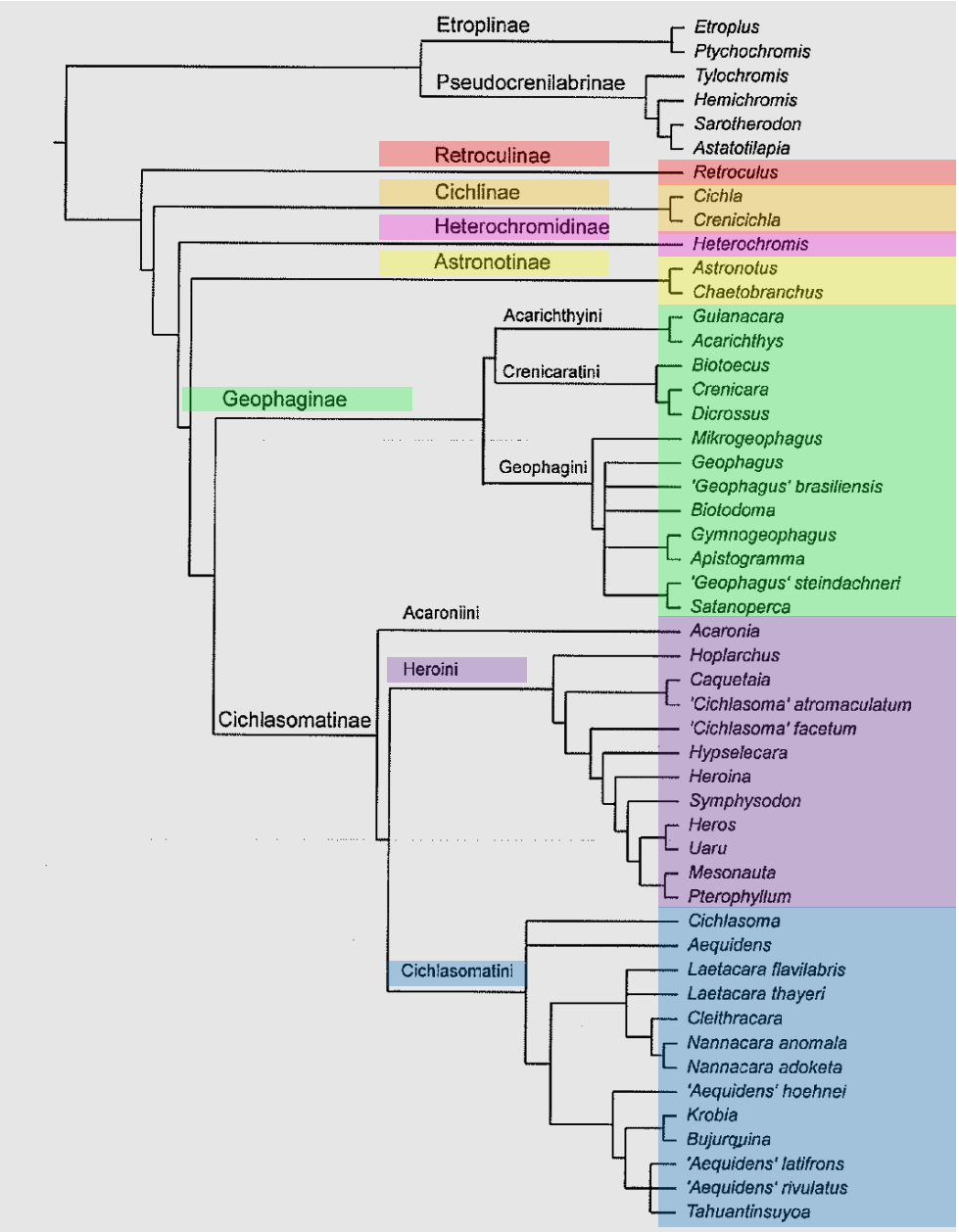


FIGURE 2.4 A strict consensus tree of 16 equally parsimonious trees as retrieved by Kullander (1998)**.** The tree was based on a parsimony analysis of 91 morphological characters and 50 terminal taxa using the Hennig86 software with a heuristic search and successive weighting.

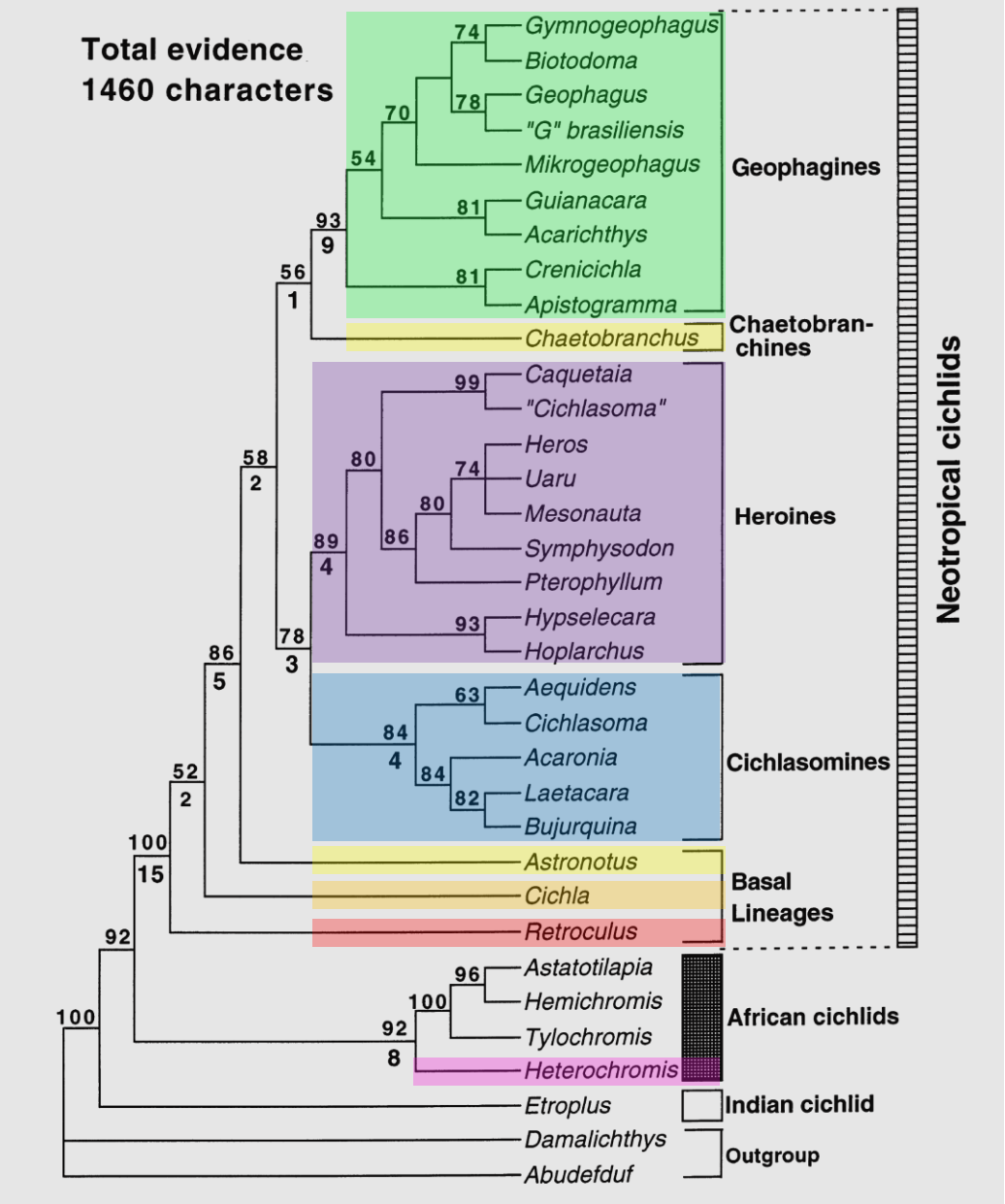


FIGURE 2.5 The strict consensus tree of three MP trees as recovered by Farias et al. (2000). The tree was constructed from a concatenated sequence matrix comprised of the three molecular markers: mitochondrial 16S rRNA, nuclear Tmo-M27 and Tmo-4C4, and 91 morphological characters (Kullander, 1998). The matrix had a total length of 1,460 nucleotide characters. Branch support values above branches corresponded to Bootstrap support values whilst those indicated below internodes corresponded to Bremer support values.

In terms of taxonomic sampling, the study by Kullander (1998) exceeded that of Farias et al. (2000) who's total evidence analysis included 34 Neotropical taxa comprising 27 genera compared to Kullander's 42 genera for a total of 47 Neotropical taxa. Following the work by Farias et al. (2000, 1999) there were no novel phylogenetic findings that were in strong disagreement with or held significant value to have brought about the formulation of a completely new phylogenetic hypothesis for the Neotropical assemblage. However, conflict and uncertainty regarding higher-level relationships among the Neotropical cichlids remained.

In the following decade, a number of new topologies were proposed for the Neotropical radiation (Farias et al., 2001; Smith et al., 2008; Sparks and Smith, 2004; Sparks, 2004). Confusion at suprageneric level among these studies was evident as the proposed topologies differed in the interrelationships of the tribes. For example, these studies differed in the inclusion of the genera *Acaronia*, *Astronotus*, *Cichla* and *Crenicichla* as well as in the proposed interrelationships of these genera. The first study in almost a decade to challenge and subsequently propose a robust alternative phylogenetic hypothesis for the Neotropical lineage was proposed by Smith et al. (2008).

2.2.1.4 Smith et al.

In their study, Smith et al. (2008) investigated the intergeneric relationships of the Neotropical cichlids in a total evidence analysis (**Figure 2.6**). The study was based on the use of primary nucleotide sequences of nine molecular markers (4 mitochondrial and 5 nuclear) in combination with the morphological data set as published by Kullander (1998). In addition, the oldest known cichlid fossil, †*Proterocara argentina*, was incorporated to infer the evolutionary history of the lineage. Consequently, Smith et al. (2008) proposed a robust phylogenetic hypothesis for the Neotropical lineage which was recovered as a monophyletic group. Of the 88 terminal taxa included in their study, 64 were Neotropical cichlids, thus ensuring that the clade was well represented. Furthermore, the study culminated in the first phylogenetic hypothesis which accurately distinguished between the seven Neotropical tribes as it is commonly acknowledged today. These tribes are Retroculini, Cichlini, Astronotini, Chaetobranchini, Geophagini, Cichlasomatini and Heroini. The latter four tribes were recovered as the two major Neotropical lineages: Chaetobranchini-Geophagini and Cichlasomatini-Heroini. Furthermore, the monogeneric tribe Astronotini was positioned between the basal lineage (Retroculini-Cichlini) and all other remaining Neotropical tribes. Hence, Astronotini was the sister-group to the clade comprising the two major Neotropical lineages. This conclusion was in agreement with the findings of Farias et al. (2000).

2.2.1.5 López-Fernández et al.

López-Fernández et al. (2010) was however of the opinion that the taxon sampling and character sets incorporated by abovementioned studies, in particular Smith et al. (2008) were insufficient to hypothesize a robust phylogeny for the Neotropical lineage. Subsequently, López-Fernández et al. (2010) constructed the largest and most comprehensive molecular phylogeny (**Figure 2.7**) at the time that had good phylogenetic resolution and could allow for in-depth investigation of speciation within the Cichlinae. The analysis was based on the combined molecular data set of three mitochondrial (Cytochrome *b*, 16S rRNA and NADH dehydrogenease subunit 4) and two nuclear (Recombinase Activator Gene 2 and ribosomal S7 intron 1) loci. Included were 57 named genera which consisted of 154 species from South and Central America. In regard to higher-level relationships among the Cichlinae, the results of this study were in agreement with the phylogenetic findings of Kullander

(1998), Farias et al. (2000) and (Sparks and Smith, 2004). However, a significant difference was the placement of the tribe Astronotini as sister clade to the major Cichlasomatini-Heroini lineage. Additionally, a close phylogenetic relationship between the two tribes Cichlini and Retroculini was recovered.

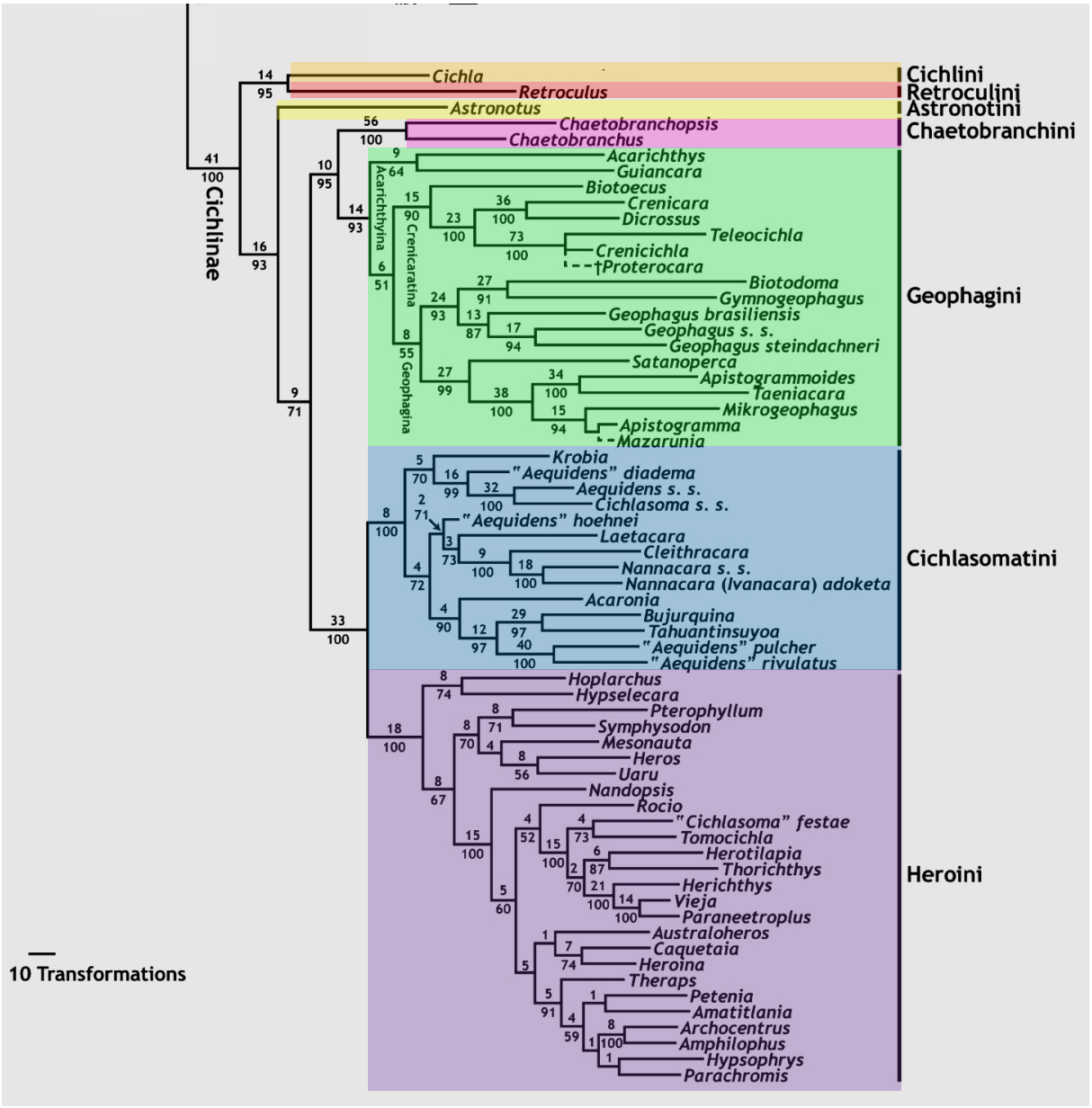
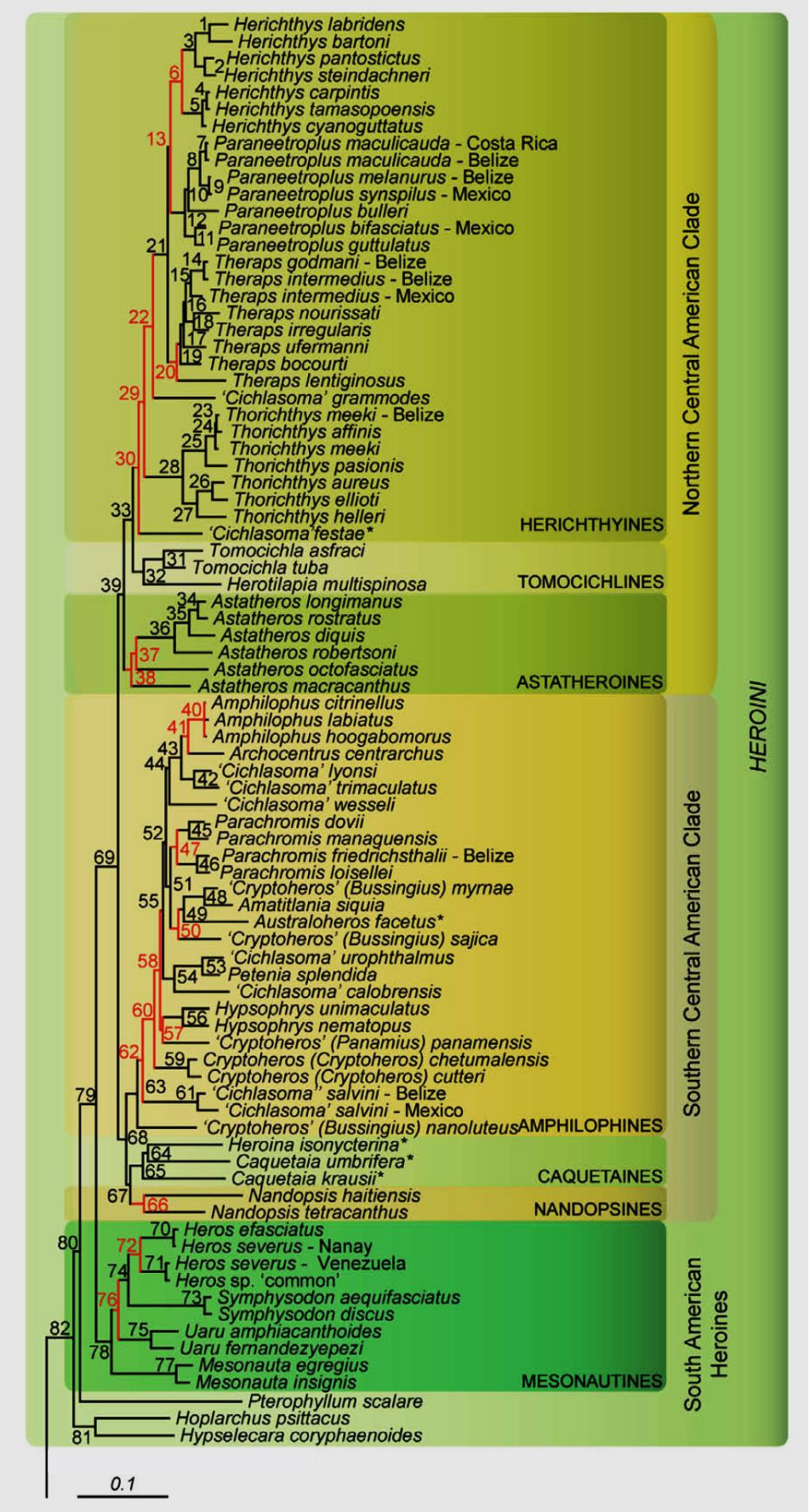


FIGURE 2.6 Parsimony tree illustrating the intergenic relationships of the subfamily Cichlinae (Smith et al., 2008). Parsimony analysis was performed on a concatenated sequence matrix consisting of four mitochondrial and 5 nuclear markers. Also incorporated were the 91 morphological characters as identified and used by Kullander 1998. In total, 88 taxa were included in the study. The taxa of the three major Cichlidae subfamilies (Etroplinae, Ptychochrominae and Pseudocrenilabrinae) and their intergenic relationships are not illustrated. Values above branches represented Bremmer support whilst those below branches represented Bootstrap support. The abbreviation s. s. means *sensu stricto* (Smith et al., 2008).



Continues on next page.

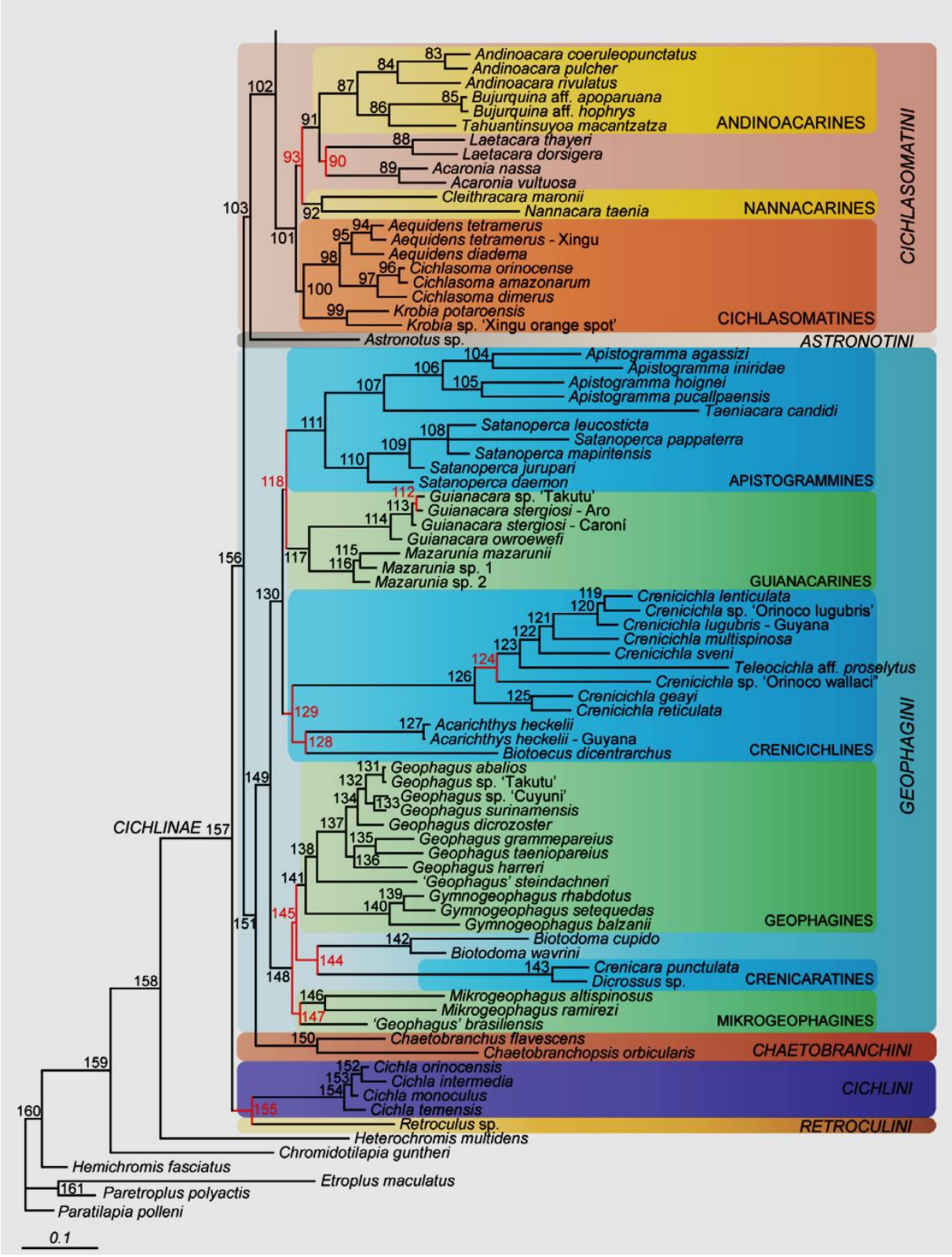


FIGURE 2.7 Phylogenetic relationships of the Neotropical lineage as recovered by López Fernández et al. (2010). The tree represents Bayesian relationships recovered in a partitioned, unlinked analysis with 12 x 10^6 replications. The partitioned data set had a total length of 3,868 nucleotide characters representative of five loci: three mitochondrial (Cyt *b*, 16S rRNA, ND4) and two nuclear (RAG2, S7). Each of the node numbers (1-161) corresponded to Bayesian posterior probability, Bootstrap and Decay index support values as summarised in Table 2 by López Fernández et al. (2010). Of interest is the 100% Bootstrap support recovered for nodes 82, 80 and 79.

The major phylogenetic findings of the abovementioned studies included the systematic and taxonomic contributions initially made by Kullander (2003, 1998) and subsequently refined by Smith et al. (2008). Also noteworthy was the consolidation of the monophyletic status of the Neotropical assemblage by almost all molecular and total evidence analyses following the morphological study by Kullander (1998). Another interesting finding was not only the inclusion of both the controversial genera *Cichla* (Stiassny, 1987) and *Retroculus* (Kullander, 1998) within the Neotropical assemblage, but also their close phylogenetic relationship and phylogenetic position as the most basal lineage of the Neotropical clade. Undoubtedly, these phylogenetic studies significantly contributed to the ultimate resolution of higher-level relationships within the Neotropical assemblage.

Currently, there is a continued interest in studying the phylogeny of the Neotropical cichlids in order to gain deeper insight into the patterns, processes and timing of their diversification (López-Fernández et al., 2013; McMahan et al., 2013). The purpose of such studies is to better understand the Neotropical lineage's complex evolution, especially in light of its assumed and speculated Gondwanan origin. Thus the task of studying the Neotropical cichlids is an ongoing initiative which is likely to continue for many years as technology improves our ability to assess the evolutionary history with greater accuracy.

2.3 TRIBE HEROINI

Kullander (1998) formally recognised the tribe Heroini which was previously referred to as the Cichlasomines Group A radiation as described by Stiassny (1991). Today, almost all of the Neotropical genera that were grouped within the informal *Cichlasoma* assemblage are recognised as valid Central American genera including the Heroine genus *Caquetaia* and Cichlasomatine genus *Acaronia*. This extends to the South American genera; *Hoplarchus*, *Heros*, *Herotilapia*, *Hypselecara*, *Mesonauta*, *Neetroplus*, *Pterophyllum*, *Symphysodon* and *Uaru* that were also representatives of the informal Cichlasoma assemblage. Although Kullander (1998) recovered the tribe Heroini as a monophyletic clade, it must be taken into account that Central American and Greater Antillean cichlids were excluded from his analyses.

2.3.1 BIOGEOGRAPHY AND PHYLOGENY

Today, Heroine cichlids are recognised to be the second most diverse tribe of Neotropical cichlids (Concheiro Pérez et al., 2007; McMahan et al., 2013; Steele and López-Fernández, 2014) and is one of a few groups of secondary freshwater fishes that is widely distributed across the American continent. According to a study by López-Fernández et al. (2010) the tribe Heroini harboured cichlid fish from two distinct geographical regions; Central and South America. This geographical range extends from southern Buenos Aires Province in Argentina to the Río Bravo/Grande river basin in Texas, USA (Concheiro Pérez et al., 2007), thus including Mesoamerica. López-Fernández et al. (2010) divided the Central American cichlids into a southern and northern clade according to their approximate

geographical locations relative to the Motagua fault. The Northern and Southern Central American clades comprised three clades each and together, constituted a total of 71 taxa representing 19 genera. Conversely, the South American clade was recovered as basal and consisted of three lineages; the closely related genera (*Hypselecara* and *Hoplarchus*), the genus *Pterophyllum* and the clade Mesonautines which comprised four genera and 10 species in total. Of interest here was the phylogenetic placement of the genus *Pterophyllum*. Previous studies (Concheiro Pérez et al., 2007; Hulsey et al., 2004), based solely on the mitochondrial Cyt *b* marker, recovered the genus as a basal taxon nested between the two tribes Geophagini and Cichlasomatini. In contrast, López-Fernández et al. (2010) not only recovered the genus *Pterophyllum* nested within the tribe Heroini, but also as one of its most basal lineages. This improved resolution can be attributed to the extended molecular data sets both in terms of genetic markers and increased taxon sampling.

2.3.2 TAXONOMIC IMPLICATIONS

The Heroine cichlids are morphologically and ecologically very diverse and have therefore been the focus of many studies reporting on ecology, behaviour, biogeography, speciation and evolutionary biology (Burress, 2015; Concheiro Pérez et al., 2007; Ilves and López-Fernández, 2014). According to McMahan et al. (2013) the tribe Heroini shows a greater species richness and an increased diversification rate in comparison to other Neotropical cichlids. It is thus not surprising that the Mesoamerican Heroines constitute approximately 25% of their environment's freshwater ichthyofauna (Concheiro Pérez et al., 2007). However, as a consequence of the tribe's diversity, phylogenetic relationships and generic allocation of most of its species are highly speculative and have remained in a constant state of flux (Kullander, 2003, 1998; López-Fernández et al., 2010; Smith et al., 2008) hence the taxonomy is unstable (Concheiro Pérez et al., 2007). This can, amongst others, be seen from the phylogenetic positions of the undescribed genera 'Cichlasoma' and 'Cryptoheros' throughout the Central American clade retrieved by López-Fernández et al. (2010). To a certain degree it also holds true for the genus *Pterophyllum*.

2.4 GENUS *PTEROPHYLLUM*

Of all the somewhat 1600 cichlid species of the fish family Cichlidae the South American genus *Pterophyllum*, or otherwise known as Angelfish, is morphologically distinct as it is shortened with considerably extended dorsal and ventral fins. Today it is commonly accepted that the genus *Pterophyllum* (Sub-family: Cichlinae; Tribe: Heroini) consists of three valid species; *Pt. scalare* (Schultze, 1823), *Pt. altum* (Pellegrin, 1903) and *Pt. leopoldi* (Gosse, 1963). Biogeographically these species are widespread throughout the Amazon River, the Orinoco River and the Essequibo River drainage basins which together, form the center of Neotropical cichlid biodiversity as these rivers harbour the majority of freshwater ichthyofauna of South America.

2.4.1 MORPHOLOGY

Morphologically, *Pterophyllum* species are laterally compressed with an extreme elongation of their dorsal and ventral fins. Other physical attributes, as noted by Kullander 1986, include a notched predorsal contour, slightly projected and prognathous lower jaw, short caudal peduncle and moderately thick lips (**Figure 2.8**). In terms of coloration, they have silvery chests with pale whitish sides and light grey foreheads. In addition they have six alternating vertical bars spanning the length of their bodies. The latter is probably the most characteristic of all their physical traits. According to Kullander (1986), the first bar is usually dark and extends from the dorsal contour, over the eye and down to the ventral contour. The second bar is more of a trace or 'shadow bar' that extends from the origin of the dorsal-fin in line with the lower lateral line. The third bar is prominent and gradually narrows extending from the 6th-7th dorsal spine along its height to just above the ventral contour. The fourth bar is another 'shadow bar' and is quite faint; stretching from the dorsal fin and almost vanishes just below the lower lateral line. The fifth bar is intense and the widest, extending from the middle of the dorsal-fin base to the anal-fin base. Finally, the sixth bar is another 'shadow bar' and the shortest as it spans the height of the caudal peduncle across the lower lateral line to the end of the anal-fin base. Another characteristic feature of the *Pterophyllum* species is a prominent bar-like mark at the base of their caudal fin which can be mistaken as the sixth and final vertical bar.

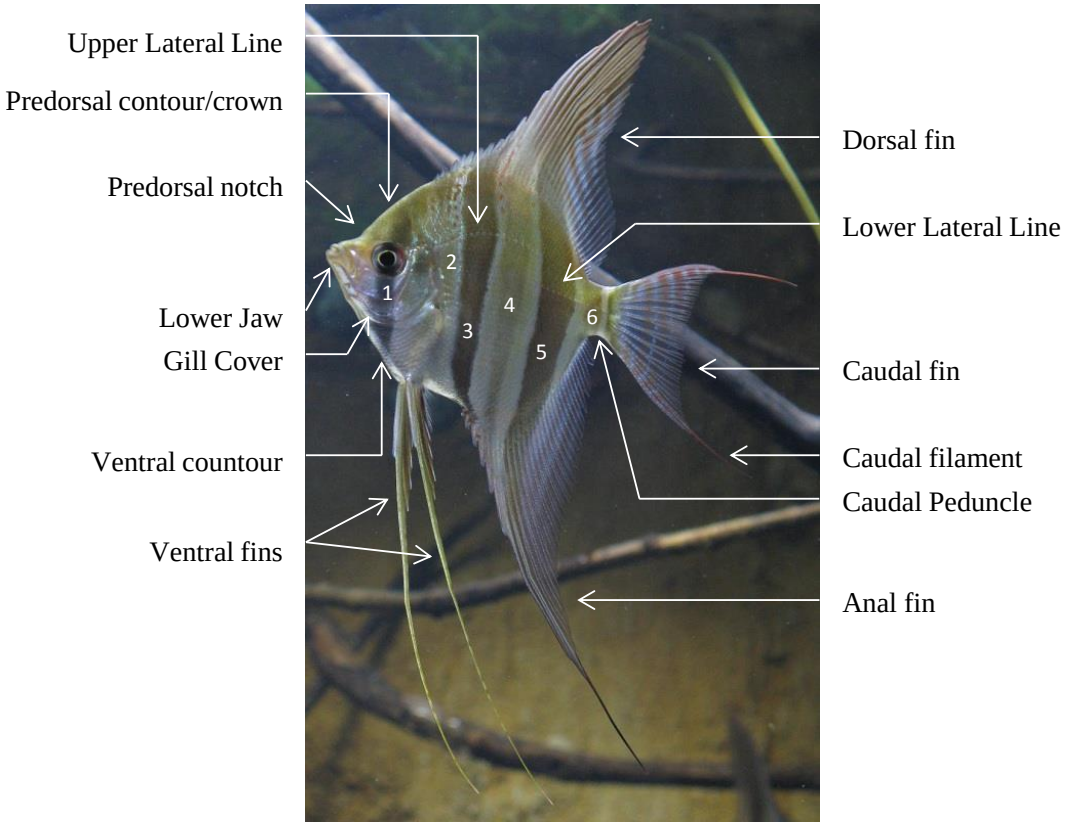


FIGURE 2.8 Physical attributes of *Pterophyllum* species. Illustrated is the basic outline of *Pterophyllum* morphology. The numbers represents their corresponding vertical bars as described under the previous heading: morphology.

2.4.2 *PTEROPHYLLUM SCALARE*

Originally described by Schultze (1823) as *Zeus scalaris*, it is currently known and accepted as *Pterophyllum scalare* according to the description by Kullander (1986). Previously, this species has been known by many different names including *Platax scalaris* (Cuvier, 1831), *Pterophyllum scalaris* (Heckel, 1840), *Plataxoïdes dumerilii* (Castelnau, 1855) and *Pterophyllum eimekei* (Ahl, 1928). The latter is sometimes still used synonymously with its current name. *Pterophyllum scalare* is the most commonly found species as it is distributed throughout the Amazon River drainage basin (including all of its tributaries), rivers of Amapá State, the Oyapock River and the Essequibo River (Kullander, 2003). Accordingly, its biogeographical range includes and spans across Brazil, Colombia, French Guiana, Guyana, Peru and Surinam. Morphologically, this species can be distinguished from its relatives based on a few distinct physical attributes. Firstly, its standard length is 7.5 cm which extends the distance from the anterior tip of the snout to the base of the caudal fin (Kullander, 1986). Secondly, it has a unique scale count ranging between 30-39 scales along the upper lateral line. Their characteristic vertical bars are naturally black in colour but the intensity of this colour can be influenced by their emotional state. Males (A) can be distinguished from females (B) by their increased numbers of rays of the ventral fin and, in some individuals, a more prominent forehead (**Figure 2.9**). Furthermore, they also have characteristic concave striations on their caudal fin that, in itself, is slightly concave with relatively short caudal filaments. Finally, they do not have a prominent spot on the dorsal-fin base at the top of the fourth vertical bar.

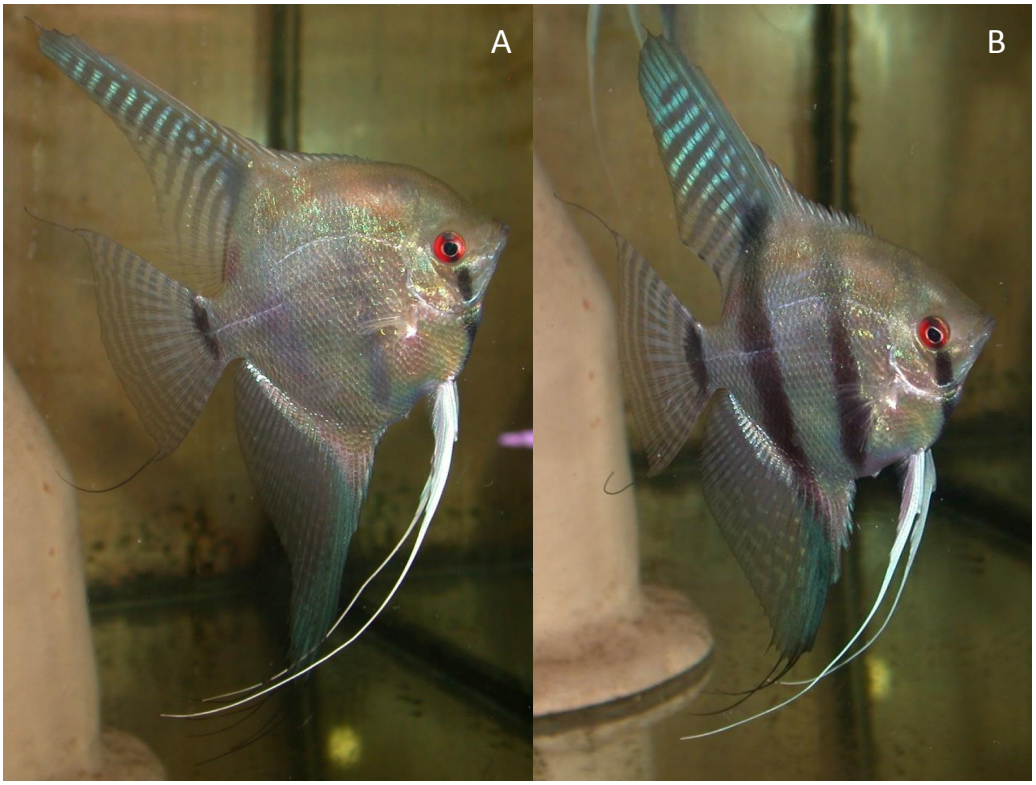


FIGURE 2.9 An image of two *Pterophyllum scalare.* Figure 2.9A shows a male whilst figure 2.9B shows a female *Pt. scalare*. Fishes are from domestic stock (D U Bellstedt). The side-to-side comparison serves to show the morphological differences between males and females.

2.4.2.1 Variants of Pterophyllum scalare

There are a few known variants of *Pt. scalare* from Peru, Lago Manacapuru and Santa Isabel. These phenotypic variants differ from the wild type *Pt. scalare* by a few characteristic morphological features. In contrast to *Pt. scalare*, Peruvian variants (**Figure 2.10**) exhibit prominent dark brown/black vertical bars alternated by light grey 'shadow bars'. The fourth vertical bar is characterised by scatter of yellow and dark brown pigmentation. The dark brown/black spots tend to accumulate around the upper lateral line and stop approximately in line with the gill cover whilst the yellow pigmentation spreads toward the ventral contour and into the anal fin. The yellow pigmentation also appears around the caudal peduncle, the base of the dorsal fin and along the predorsal contour as a prominent yellowish/brownish 'stain'. A prominent dark spot is also visible right above the fourth vertical bar at the base of the dorsal fin. In the trade, these variants are referred to as 'Peru altums' and although these are distinctive members of *Pt. scalare*, the use of the name 'altum' is aimed at misleading the buyer and obtaining a higher price for the fishes.

FIGURE 2.10 *Pterophyllum scalare*. Illustrated are examples of Peruvian variants of *Pt. scalare* caught in the Solimões River within Peru. Photo by Joost Boumans (Boumans, 2015)

The Lago Manacapuru variants (**Figure 2.11**) can be distinguished by their dark orange/reddish coloration of their dorsal fins that spread from above the lateral line along the predorsal contour. This coloration is darkest at the base of the dorsal fin, at the top of the fourth vertical bar, and then gradually fades toward the snout area including the eye and lower jaw. This variant is popular in the trade because of its red colour.

FIGURE 2.11 *Pterophyllum scalare*. Illustrated is an example of an immature Lago Manacapuru variant of *Pt. scalare* caught in Amazon River within the vicinity of Lago Manacapuru. Photo supplied by D U Bellstedt.

Recently a variant has been captured from the vicinity of Santa Isabel on the upper Rio Negro (**Figure 2.12**). These fish appear to share a few anatomical and morphological features with both *Pt. scalare* and *Pt. altum*. Anatomically, they are more triangular than round; a feature they share with *Pt. altum*. Furthermore, they have characteristic red pigmentation within the region above their lateral line which extends into the dorsal fin. The latter is another characteristic feature they share with *Pt. altum*. In comparison with *Pt. scalare*, their vertical bars, even though dependent on their mood, are also naturally light grey in colour. Furthermore, their caudal fins also show a pattern of concave striations in addition to the similarities in length of their dorsal, anal and ventral fins. Unique to the Santa Isabel variants is their characteristic yellow coloration in their dorsal, anal and caudal fins as well as their crown area along the predorsal contour and within the region of the eye.

FIGURE 2.12 *Pterophyllum scalare*. Illustrated are examples of Santa Isabel variants of *Pt. scalare* caught in the upper Rio Negro in the vicinity of the town of Santa Isabel. Photo supplied by Joost Boumans.

2.4.3 *PTEROPHYLLUM ALTUM*

Pterophyllum altum was described by Pellegrin in 1903 from the Rio Atabapo, an upper tributary of the Rio Orinoco. *Pterophyllum altum* is predominantly distributed throughout the Orinoco River basin, more specifically the tributaries of the Upper Orinoco River to Puerto Ayacucho (Kullander, 2003). However, there is considerable controversy about its occurrence in the upper Rio Negro tributary of the Amazon River which in turn is linked to the Upper Rio Orinoco by means of the Casiquiare Canal. Some authors (Bleher, 2012) claim these to be *Pt. scalare* variants or hybrids of *Pt. scalare* and *Pt. altum*. Thus, *Pt. altum* is confined within the boundaries of Colombia and Venezuela and possibly Brazil. Morphologically, *Pt. altum* possesses a number of unique physical attributes. First and foremost, it has characteristically elongated dorsal- and ventral-fins reaching up to 40 cm in height making it the tallest of the three species, if not the tallest of all cichlids (Burress, 2015). It is also more triangular and deep-bodied than its sister species (Kullander, 1986) and has a maximum standard length of 6.5 cm (Kullander, 2003). In addition, its characteristic vertical bars are naturally black and much wider than those of its relatives (Kullander, 1986). Furthermore, it has a more prominent predorsal notch whilst a prominent spot on the dorsal-fin base is absent. Another characteristic feature includes the red pigmentation concentrated within the region directly above the upper lateral line extending toward the dorsal contour and into the dorsal fin. According to Kullander (1986) the *Pt. altum* has a scale count of 46-48 scales on the lateral line which is the highest of the three species.

The latter is generally regarded as the most accurate means to distinguish it from *Pt. scalare*. Examples of *Pt. altum* from its type locality i.e. the Rio Atabapo as well as the Rio Inirida and the upper Rio Negro are shown in **Figures 2.13**, **2.14** and **2.15** respectively.

FIGURE 2.13 ***Pterophyllum altum.*** Wild type *Pt. altum* caught in the Rio Atabapo as kept in the facility of the ornamental fish breeder Simon Forkel (Forkel, 2015). Photo supplied by D U Bellstedt.

FIGURE 2.14 ***Pterophyllum altum.*** Wild type fishes caught in the Rio Inirida (front) and the Rio Atabapo (back) as kept in the facility of the ornamental fish breeder Simon Forkel (Forkel, 2015). Photo supplied by D U Bellstedt.

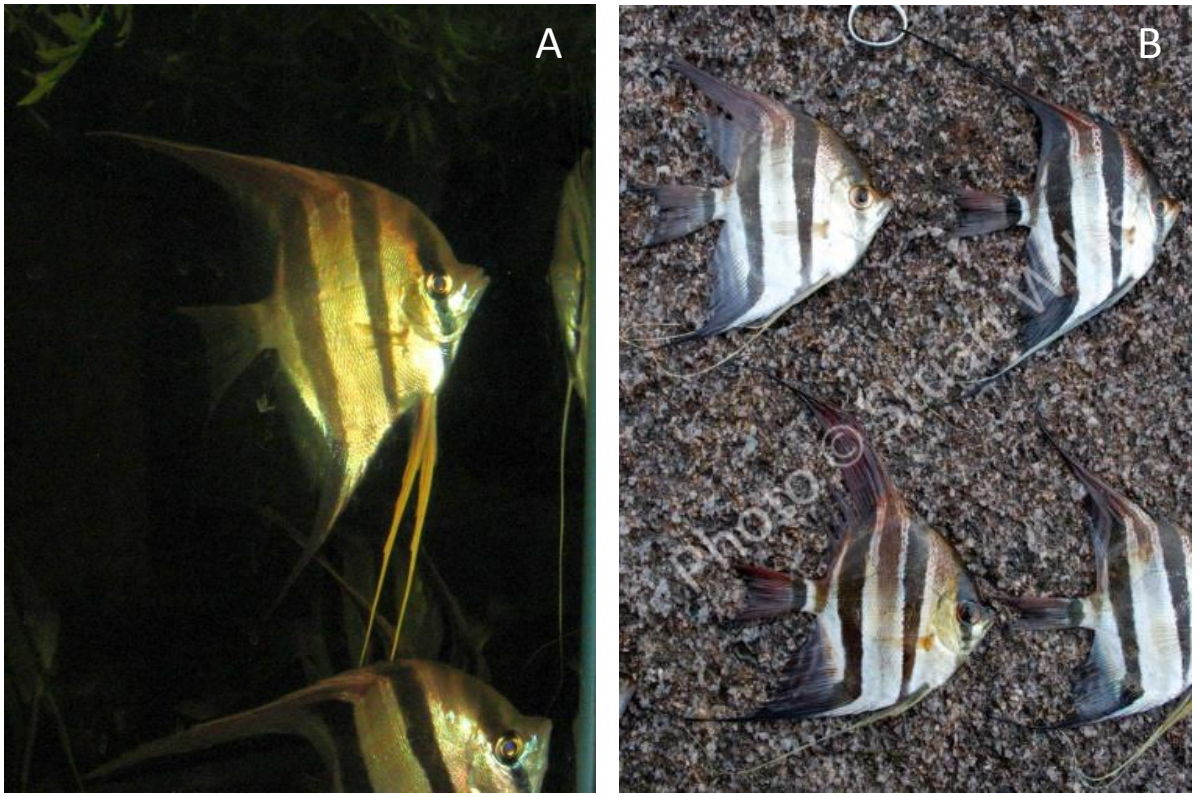


FIGURE 2.15 *Pterophyllum altum*. Fishes caught in the Rio Negro (A) and the Rio Uaupés (B). The latter is a tributary of the upper Rio Negro. Photo provided by Mr Edgar Ruiz and those from the Rio Uaupés by Stewart C Willis (Willis, 2015).

2.4.4 *PTEROPHYLLUM LEOPOLDI*

Pterophyllum leopoldi (**Figure 2.16**) was first described by Gosse in 1963 as *Plataxoides leopoldi* but once its relatedness to *Pt. scalare* and *Pt. altum* was confirmed, it was formally described as *P. dumerilii* with *Pterophyllum leopoldi* a synonym (Schultz, 1967). This species is mainly confined to the Rupununi River of the Essequibo River basin in Guyana and along the Solimões-Amazon River between Lago Manacapuru and Santarém in Brazil (Kullander, 2003). *Pterophyllum leopoldi* differs from its relatives in a number of ways. Not only is it relatively slender, large-scaled and short-finned, it has fewer scales along the lateral line (27-29) and a truncated caudal fin with marginal filamentous extensions (Kullander, 1986). It has a maximum standard length of 5 cm and hence is the smallest of the three *Pterophyllum* species (Kullander, 2003). However, it can best be distinguished from its relatives by its characteristic straight (unnotched) predorsal contour and black spot on the dorsal-fin base above the fourth vertical bar (Kullander, 1986). According to Kullander 1986, the morphological characteristics of *Pt. leopoldi* are such that it can be considered the 'morpological intermediate' between the neighbouring Herioni genus *Mesonauta* and *Pterophyllum.*

FIGURE 2.16 *Pterophyllum leopoldi*. This species is morphologically very different from its relatives as it has a characteristically short but elongated body, a straight and unnotched predorsal contour, prominent dorsal spot at the base of the dorsal fin and significantly shorter fins. Photo by Gert Blank, the Netherlands.

2.4.5 MORPHOLOGICAL VARIATION WITHIN GENUS *PTEROPHYLLUM*

Morphological differences between the three species (*Pt. leopoldi*, *Pt. altum* and *Pt. scalare*) of the genus *Pterophyllum* and their phenotypic variants are listed in **Table 2.1**.

TABLE 2.1 A summary of the morphological differences between the three *Pterophyllum* species (highlighted in blue) and their phenotypic variants (highlighted in purple). The four most distinctive morphological features are their lateral line scale count, presence of a predorsal notch, the curvature of their caudal fin margin and the presence of a prominent black spot at the base of the dorsal fin.

Species	Lateral line scale count	Predorsal notch	Caudal fin margin	Prominent black spot
Pt. leopoldi	27-29	No	Concave	Yes
Pt. scalare	30-39	Yes, but variable	Concave	No
Pt. altum	46-48	Yes	Convex	No
Santa Isabel	ND	Yes	Convex	No
Lago Manacapuru	ND	Yes	Concave	No
Peru	ND	Yes	Concave	No

2.4.6 PHYLOGENY

Despite the comprehensive knowledge of the genus *Pterophyllum*, it has been underrepresented at species-level in the majority of previously published phylogenetic studies of the Cichlinae and/or Heroini. Furthermore, the phylogenetic relationship of the genus with its neighbouring South American Heroine cichlids remains controversial. In her review of cichlid interfamilial relationships, Stiassny (1991) recovered a close phylogenetic relationship between *Pterophyllum* and *Symphysodon* nested within her informal Cichlasomines Group A. On the other hand Kullander (1998) recovered a close phylogenetic relationship between *Pterophyllum* and *Mesonauta* within his formally recognised tribe Heroini. Although both these morphological studies correctly placed *Pterophyllum* in the South American Heroine cichlids (López-Fernández et al. 2010), consensus regarding its generic placement within the Neotropical assemblage had not been reached.

Previous molecular and combined analyses (total evidence) appeared to be unable to resolve the controversy either. Farias et al. (1999) recovered the genus as the basal lineage of a mainly South American Heroine clade. However, in their subsequent study (Farias et al., 2000), they recovered *Pterophyllum* nested between the basal lineage *Hypselecara-Hoplarchus* and a clade containing all other South American Heroine genera. Interestingly, Farias et al. (2001) recovered *Pterophyllum* as the most basal lineage of the tribe Heroini which consisted of both Central and South American Heroine genera. Based on the results of a strict consensus tree, Sparks (2004) recovered *Pterophyllum* nested within an 'inner' Heroini clade consisting of the genera *Symphysodon*, *Parachromis*, *Archocentrus* and *Caquetaia*. In addition, Concheiro Pérez et al. (2007) recovered the genus *Pterophyllum* as sister-group of the ancestral Cichlasomatini-Heroini lineage.

It appears that the phylogenetic position of the genus *Pterophyllum* within the tribe Heroini as retrieved in more recently published phylogenies are in agreement with those previously hypothesised For example, the phylogenetic position of the genus *Pterophyllum* as retrieved by Smith et al. (2008) is congruent with the placement of Stiassny (1991) whilst the phylogenetic position of the genus *Pterophyllum* retrieved by López-Fernández et al. (2010) and Říčan et al. (2013) are congruent with that of Farias et al. (2001). However, despite this, the position of the genus *Pterophyllum* within the tribe Heroini remains unclear. Accordingly, by resolving the phylogenetic relationships of the three species of *Pterophyllum* with confidence in relation to its related genera in the Heroini will undoubtedly contribute to establishing the correct taxonomic position of the genus.

To date, Meliciano (2009) conducted a phylogeographic study of the genus *Pterophyllum* based on the use of the mitochondrial Cyt *b* gene and geometric morphometric data. Taxon sampling was thorough with 328 specimens collected from 14 locations in the major tributaries of the Amazon River. Two hundred and forty-four taxa were included in the analysis which consisted of 5 *Pt. leopoldi*, 24 *Pt. altum* and 215 *Pt. scalare* taxa. An interesting conclusion was drawn from her study. Based on

morphometric measurements, fishes from Santa Isabel and Boa Vista, in the upper Rio Negro, appeared to be phenotypically related to *Pt. altum*. According to the molecular analysis, based on the mitochondrial Cyt *b* gene, it was evident that these populations either had a *Pt. scalare* copy of the Cyt *b* gene or a distinctly different copy of the gene which was assumed to be *Pt. altum*. The results thus strongly suggested that *Pt. scalare* and *Pt. altum* populations share the same habitat. However, the sampling locations for *Pt. altum* were restricted to the areas of Santa Isabel and Boa Vista. Hence, *Pt. altum* from its type locality i.e. the Rio Atabapo and tributaries of the Upper Rio Orinoco and the Rio Casiquiare were excluded. Consequently, the designation of these sequences by Meliciano (2009) as representative of *Pt. altum* could not be made unequivocally. Furthermore, her outgroup taxa were limited to only *Symphysodon discus* thereby potentially weakening the tree topology (Vandamme 2009).

2.5 CICHLID BIOGEOGRAPHY

To fully comprehend the integration of the phylogeny and biogeographical history of Cichlidae (Cichlinae: Heroini: Genus *Pterophyllum*), it is best to start this discussion with the ancient landmass called Gondwana. Gondwana was one of two supercontinents which existed throughout the Cambrian to Early Jurassic periods from approximately 510 to 180 Ma ago (Rogers and Santosh, 2004). This supercontinent included the majority of today's landmasses such as Antarctica, South America, Africa, Madagascar, Australia and India. East Gondwana consisted of modern Antarctica, Madagascar, India and Australia whilst the conjoined African and South American landmasses were known as West Gondwana. During the Early Jurassic (~184 Ma ago), East Gondwana started to separate from West Gondwana. A time period of approximately 50 Ma ago elapsed until the Early Cretaceous (~130 Ma ago) when the South Atlantic Ocean began to open (Blenkinsopand and Moore, 2003). This resulted in the split of West Gondwana characterised by the slow westward drift of the South American landmass away from the African landmass. In contrast, East Gondwana only started to separate at around 120 Ma ago which saw India migrate to the Northern Hemisphere whilst Madagascar remained south-east of the African continent. As a consequence of this continental drift the Atlantic, Pacific and Indian Oceans were allowed to ensue at around 110 Ma ago.

Accordingly there are two main hypotheses for the continental distribution of modern cichlids. One hypothesis states that the family Cichlidae have an Early Cretaceous (130 Ma ago) (Genner et al., 2007) or possibly Late Cretaceous (96-67 Ma ago) origin (Farias et al., 1999; McMahan et al., 2013). Many phylogenetic studies based on morphological (Kullander, 1998; Stiassny, 1991, 1987), molecular (Farias et al., 1999; Karl and Streelman, 1997; Sparks, 2004; Streelman et al., 1998; Zardoya et al., 1996a) and combined data sets (Farias et al., 2000) are in support of drift-vicariance events as an explanation for the modern continental distribution of Cichlidae. Consensus reached in this regard was based on the agreement between the timing of the origin of the family Cichlidae and the fragmentation of Gondwana. The general assumption thus made is that cichlids were inhabitants of

Gondwana's freshwater resources prior to the start of its disintegration. However, with the subsequent emergence of marine waters (South Atlantic and Indian oceans) further dispersal of cichlids was limited as the oceans became physical barriers. Consequently, the four sub-families each evolved separately under the influence of biotic factors unique to each continent.

According to Murray (2001) this was based on two unfounded assumptions: firstly, that the distribution of cichlids conforms to the separation of the ancient continent of Gondwana and secondly, that cichlids are incapable of dispersal through the sea i.e. saline water. Furthermore, Murray (2001) was of the opinion that the aforementioned hypothesis lacks one crucial piece of supporting evidence i.e. cichlid fossils true to the Cretaceous have not yet been discovered. Based on her knowledge and incorporation of the oldest known cichlid fossils, Murray (2001) is of the opinion that the family Cichlidae has an Early Tertiary (65 Ma ago) origin. In contrast to Gondwana-based-hypotheses, Murray's reconstruction of cichlid biogeography acknowledge that, as secondary freshwater fishes, cichlids had the ability to tolerate marine waters (Murray, 2001). Furthermore, it takes into consideration the biogeographical distribution of modern and fossil cichlids as well as their phylogenetic relationships (Murray, 2001). Therefore her reconstruction is consistent with the fossil record. As such, Murray (2001) is of the opinion that an Early Cretaceous origin of more than 130 Ma ago is excessive and that an early Tertiary or even Palaeocene dating is more credible. This was supported by two more recent independent studies (Friedman et al., 2013; McMahan et al., 2013) in which the dates of fossils of outgroups and the dates of cichlid fossils respectively were used to date phylogenies which gave dates for the divergence of the Cichlidae radiation as 81 Ma ago (67-96 Ma ago; 95% HPD) and 64.9 Ma ago (57.3-76.0 Ma ago; 95% HPD) respectively.

2.6 SOUTH AMERICAN GEOGRAPHY

The South American continent is one of the distinct biogeographical regions of the world and constitutes the majority of the Neotropic ecozone (**Figure 2.17**). For the most part, South America is located in the Southern Hemisphere except for five of its countries - Colombia, Venezuela, Guyana, Surinam and French Guiana which are located in the Northern Hemisphere. It is bordered to the west by the Pacific Ocean and by the Atlantic Ocean to the north and east. The continent has a total area of 17, 840, 000 km^2 of which the western territory is dominated by the Andes mountain range whilst the eastern terrain is dominated by the South American Platform (Albert and Reis, 2011). Both the Guiana and Brazilian Shields, characterised by highland regions and large, relatively flat lowlands, are embedded in the South American Platform (Albert and Reis, 2011). Of geographical significance are the continents' two major river drainages - the Amazon and Orinoco River Basins. Combined, these two drainages make up almost 45% of the continents total area and are the main sources of its very diverse freshwater ichthyofauna.

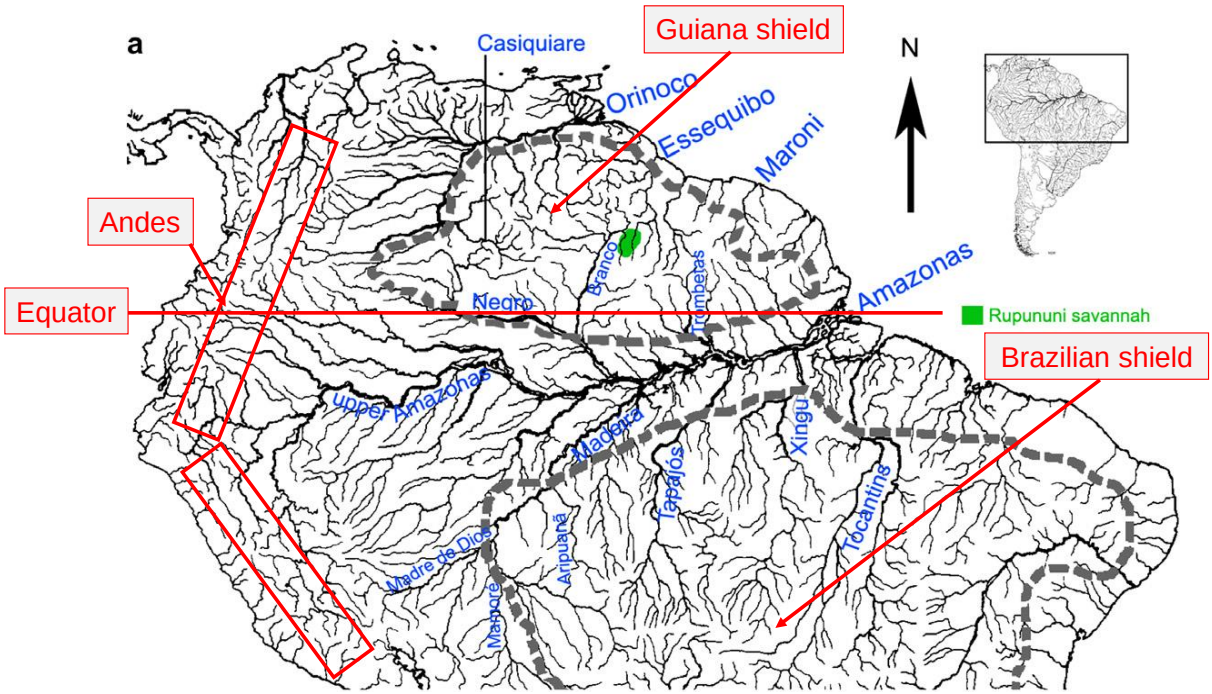


FIGURE 2.17 Major drainage systems of northern South American continent. Names of some major river systems are annotated outside of the South American contour. Names, marked in blue, within the South American contour represent some of the major tributaries of the Amazon River. Also annotated are the Equator, Andes mountain range and both the Guiana and Brazilian Shields. The latter are enclosed with grey dashed lines (Willis et al., 2007).

From a geological perspective, the Amazon and Orinoco drainage basins are considered relatively young when taking into account the age of the freshwater lineages they harbour (Albert and Reis, 2011). It is believed that the Amazon and Orinoco basins developed during the Cretaceous (~130 Ma ago) and adopted their modern configurations some 23-7 Ma ago during the Neogene period (Albert and Reis, 2011). Based on the Late Cretaceous to early Eocene (74-54 Ma ago) origin proposed for the subfamily Cichlinae by McMahan et al. (2013), it appears that the main geological features of the Neotropics, which include the Andes mountain range and the Sub-Andean Foreland, have guided the development of the major South American drainage basins which resulted in the modern biogeographical distribution of Neotropical cichlids (Albert and Reis, 2011). Accordingly, Lévêque et al. (2007) suggested that the patterns of the modern biogeographical distribution of Neotropical freshwater fish populations, confined to these drainage systems, could yield insight into past continental changes. Thus, having a firm understanding of Amazonian landscape and how it developed over time, together with phylogenetic and biogeographic information of Neotropical freshwater fish, could yield insight into their evolutionary history.

2.7 AMAZONIAN LANDSCAPE AND DRAINAGE PATTERNS THROUGH TIME

The key to understanding modern Amazonian landscape is revealed through Miocene plate-tectonic readjustments in the north-eastern Andes. In short, these tectonic events initiated the reorganisation of the former east to west drainage patterns into the modern-day west to east Andean drainage system (Hoorn et al., 1995). The Andes is a terrestrial mountain range of consecutive highlands which extends

a distance of over 7,000 km along the western coast of the South American continent (Albert and Reis, 2011). It encompasses two parallel mountain ranges; the Eastern and Western Cordilleras. The Andes can be divided into three main sections based on the timing of concessive tectonic uplifts and their locations, namely the Southern, the Central, and the Northern Andes (Albert and Reis, 2011). Of geographical significance are the north-eastern Andes which arose through a series of at least six phases of uplift and intermittent tectonic stability during the Late Cretaceous to Pleistocene (Albert and Reis, 2011; Hoorn et al., 1995). Its main uplift was the Eastern Cordillera which occurred throughout the Neogene-Quaternary period. The uplift can be described in three distinct time intervals; late Oligocene to early middle Miocene, late middle Miocene and late Miocene to Pleistocene (Hoorn et al., 1995).

During the late Oligocene to early middle Miocene (**Figure 2.18**), pan-Amazonia was characterised by the predominant paleo-Orinoco river system, also described as the Pebas system (Albert and Reis, 2011; Hoorn et al., 2010). This ancient river system had a few tributaries which drained the Central and emergent north-eastern Andes to its west as well as the Guiana and Brazilian shields to its east (Albert and Reis, 2011; Hoorn and Wesselingh, 2010; Hoorn et al., 1995). The Pebas system had a northward course through the Andean foreland basins and deposited its sediment into the Llanos basin which was open to marine waters of the East Venezuela basin (Albert and Reis, 2011; Hoorn et al., 2010). At this time the Northern Andes, more specifically the Eastern Cordillera, was in its early stages of development and was not yet prominent (Hoorn et al., 2010, 1995). Also of significance was the Purus Arch; a major drainage divide between eastward and westward flowing rivers located in central-eastern Amazonia (Albert and Reis, 2011; Hoorn and Wesselingh, 2010). A Maracaibo-Falcon marine connection was also evident (Albert and Reis, 2011; Hoorn et al., 2010, 1995). Although phases of global sea level rise were reported by Hoorn et al. (1995), uncertainty remains whether a physical connection existed between Lake Maracaibo and the Llanos basin (Albert and Reis, 2011). Thus the subject of marine incursions, during the Miocene, remains controversial (Hoorn and Wesselingh, 2010).

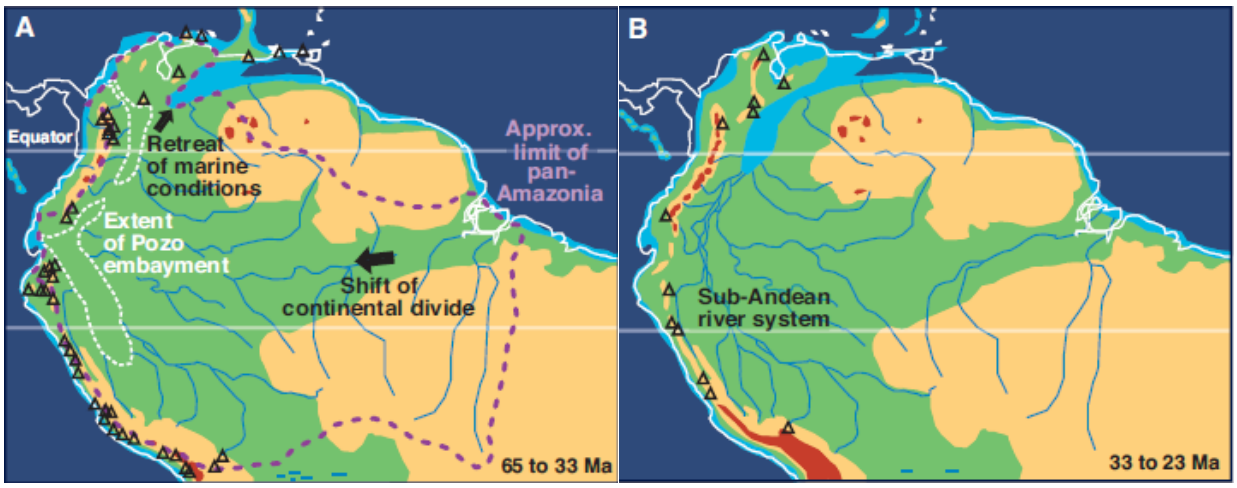


FIGURE 2.18 Paleocene to early middle Miocene. Figures A and B as published by Hoorn et al. (2010).

For the majority of late middle Miocene (**Figure 2.19**), the paleo-Orinoco River or Pebas system was the predominant waterbody and continued its course north to the Llanos basin. However, due to continued uplift of the Northern Andes the Eastern Cordillera became prominent and subsequently initiated significant changes in drainage patterns (Hoorn et al., 1995). Consequently, this time period saw the Pebas system subside and lead to the development of the predominant paleo-Amazon River. The latter was a system of rivers and lakes containing sediments with a flow in an eastward direction which thus became the precursor of the present-day Amazon River (Hoorn et al., 1995). Although receded, the paleo-Orinoco River was still partly connected to the paleo-Amazon River and together these two river systems thus formed a meandering fluvial sytem known as the paleo-Amazon-Orinoco River basin which engulfed the majority of the western Amazonian territory (Albert and Reis, 2011; Hoorn et al., 1995).

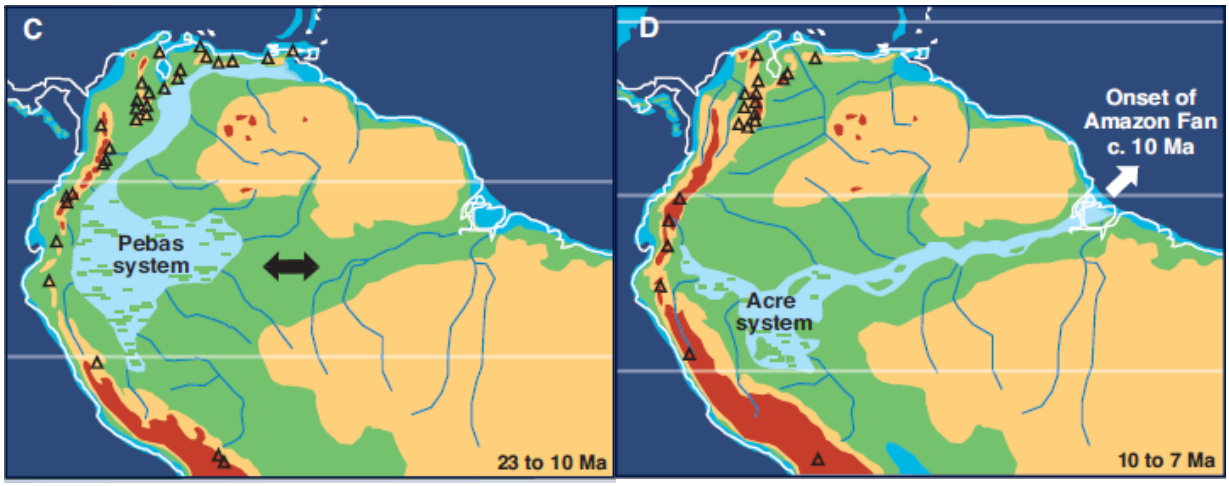


FIGURE 2.19 Late Middle Miocene. Figures C and D as published by Hoorn et al. (2010).

The period between the late Miocene and early Pleistocene (**Figure 2.20**) marked the most dynamic episode of accelerated uplift of the north-eastern Andes (Hoorn and Wesselingh, 2010; Hoorn et al., 2010). Accordingly it saw the Andes attain their present configuration of which the uplift of the Eastern Cordillera and Mérida Andes was most significant (Hoorn and Wesselingh, 2010; Hoorn et al., 2010, 1995). Furthermore, the uplift of the Vaupes-Arch (Vaupes-Guaviare region) separated the paleo-Amazon-Orinoco River into two separate drainages (Albert and Reis, 2011; Hoorn and Wesselingh, 2010; Willis et al., 2010; Winemiller et al., 2008). With the final elevation of the Mérida Andes, the Orinoco drainage was forced to shift east. At the same time the Llanos basin was separated from the Amazon system to its west, the Maracaibo/Falcon region to its north and from its marine connection with the East Venezuela basin to its east. At approximately the same time the paleo-Amazon River broke through its eastern barrier, the Purus Arch, which initiated its transformation into a transcontinental drainage system by reaching the Atlantic Ocean (Hoorn et al., 2010, 1995).

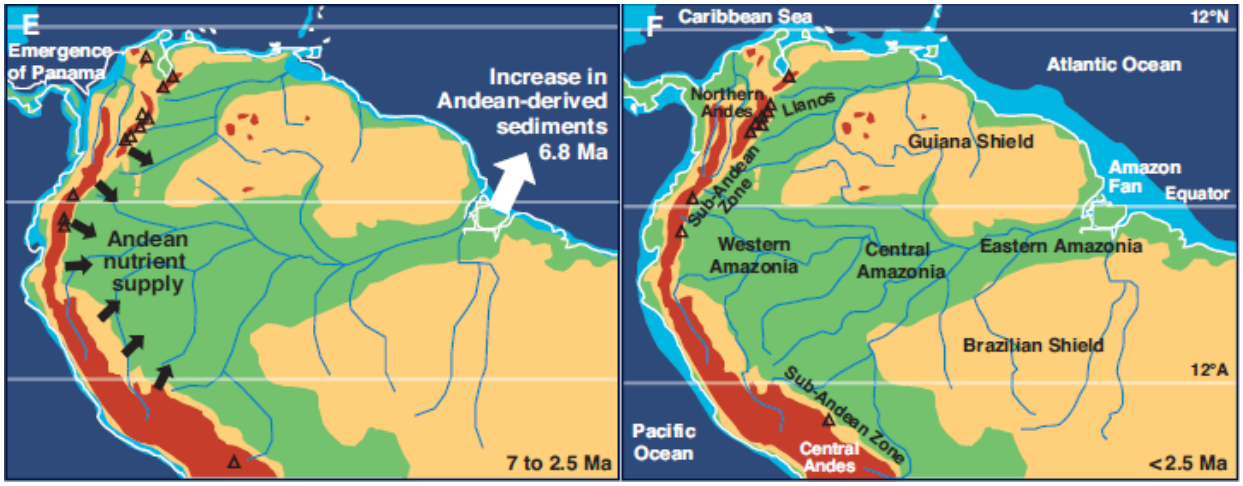


FIGURE 2.20 Late Miocene to Pleistocene. Figures E and F as published by Hoorn et al. (2010).

2.8 PRESENT DAY AMAZON AND ORINOCO BASINS

The Amazon drainage basin is the world's largest fluvial hydrographic system with a total surface area of 6,915,000 km^2 (Albert and Reis, 2011; Winemiller et al., 2008). This means that it covers about 40% of the South American continent. Its source is the headwaters of the Ucayali River in the Andes, Peru, from which it runs eastward for a total length of 6,400 km through the countries of Brazil and Colombia to where its flow is discharged through a broad estuary into the Atlantic Ocean (Albert and Reis, 2011; Hoorn and Wesselingh, 2010). To avoid confusion, the basin's mainstem is referred to as the Amazon River, although geographically it is known by two names. From its origin to where it conjoins with its largest tributary, the Negro River, in Manaus, it is known as the Solimões River. Between Manaus and its mouth in the Atlantic, it is referred to as the Amazonas or Amazon River (Albert and Reis, 2011; Kullander, 1986). The river's major tributaries include the Negro, Japurá and Trombetas on its north bank in addition to the Purus, Madeira, Tapajos and Xingu on its south bank (Albert and Reis, 2011). Its water chemistry is highly varied and in general, it appears that the rivers draining the Guiana and Brazilian shields have black- (low pH, low conductivity, high organic content) to clear- (high transparency) waters (Albert and Reis, 2011). In contrast, the rivers draining the Andes tend to have sedimented so-called white waters with neutral pH and low transparency (Albert and Reis, 2011).

The Orinoco basin is considered the third largest in the world with a total surface area of 983,000 km^2 and thus covers about 5% of the South American continent (Albert and Reis, 2011; Winemiller et al., 2008). The Parima Mountains, located on the border of Venezuela and Brazil, serve as the source of the Orinoco River. From here it extends a distance of about 1,500 km to the north and then northeast along the northern margin of the Guiana Shield to reach its mouth, the Delta Amacuro, on the Caribbean coast of Venezuela (Albert and Reis, 2011). The river can be divided into two main stretches referred to as the Upper and Lower Orinoco. The Guaviare River is considered the major tributary of the Orinoco basin based on its superior geographical features when compared to those of the Upper and Lower Orinoco (Albert and Reis, 2011). The Orinoco River has many tributaries along

its length and these include, but are not limited to, the Atabapo, Caroní, Capanaparo, Guaviare, Inírida, Meta and Ventuari. Of all the major Upper Orinoco tributaries, the Guaviare and Meta are known to be whitewater rivers due to the high volumes of sediment that they receive from the Guiana Shield. Hence, these rivers have slightly acid pH and a moderate concentration of solutes (Winemiller et al., 2008). In contrast, tributaries such as the Ocamo, Ventuari, Atabapo, Inirida and Vichada consist mostly of clearwater with extremely acidic waters and extremely low concentrations of solutes (sodium, magnesium, calcium and potassium) resulting in water with very low conductivity. A peculiarity of the Orinoco River system is the Casiquiare Canal which is a outflow of the Orinoco River and is considered the largest river to connect the Orinoco and Amazon drainage basins via bifurcation (Albert and Reis, 2011).

2.9 THE CASIQUIARE CANAL

The Casiquiare Canal originated as a bifurcation of the Upper Orinoco's main channel as a result of Andean erosion and subsequent sediment accumulation within the Llanos floodplain following the uplift of the Vaupes Arch (**Figure 2.21**) (Willis et al., 2010; Winemiller et al., 2008). Thereby, it is a large natural waterway that formed a permanent aquatic connection between the Orinoco and Amazon drainage basins at presumably less than eight Ma ago (Winemiller et al., 2008). The precise timing of the connection between the Casiquiare Canal and the Upper Orinoco remains unknown (Winemiller et al., 2008). The headwaters of the Upper Orinoco serve as the Casiquiare's source from which it extends about 320 km in a general south-westerly direction, across southern Venezuela, to reach its mouth in the upper Rio Negro near the town of San Carlos (Albert and Reis, 2011). In addition to its characteristic winding course and several rapids, it has many large tributaries (Winemiller et al., 2008). In terms of its chemical composition, the Casiquiare Canal represents a hydrochemical/environmental gradient between the clearwaters of the Upper Orinoco, at its origin, and the blackwaters of the upper Negro River, at its mouth (Winemiller et al., 2008). Consequently, questions surrounding its function as a free dispersal corridor or as a partial barrier to the interchange of fish species between the Orinoco and Amazon drainage basins have recently been investigated (Albert and Reis, 2011; Willis et al., 2010; Winemiller et al., 2008). Conclusions of these studies are all in agreement on two important points.

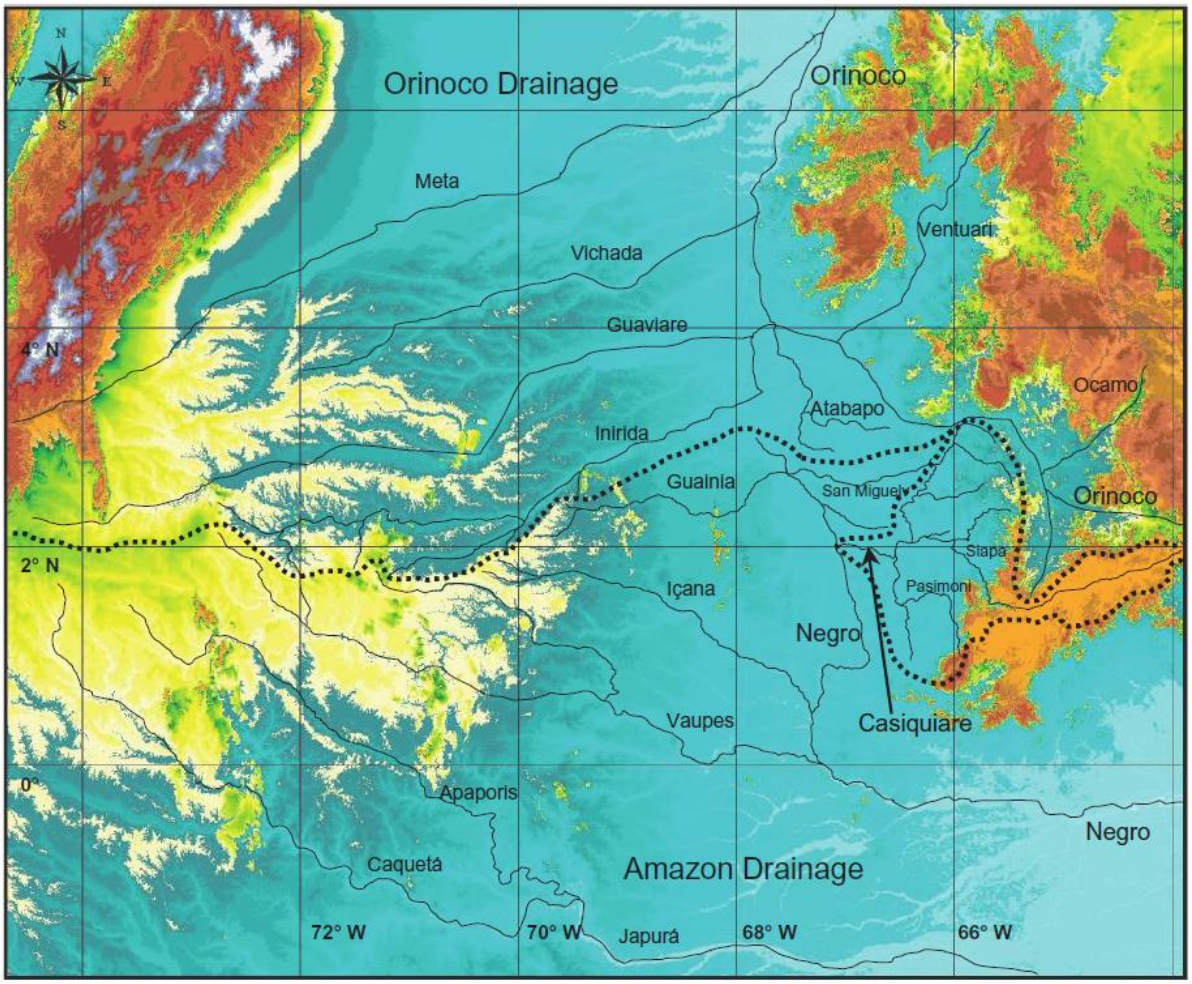


FIGURE 2.21 Digital elevation map for the region of the Vaupes Arch in southeastern Colombia and southwestern Venezuela. Elevation ranges from 25–50 m above sea level (light blue) to 2,500–2,750 m above sea level (dark red). Major river courses are overlain as thin black lines. Watershed divides for the Amazon, Orinoco, and Casiquiare basins appear as dotted lines (Winemiller and Willis, 2011).

Firstly, the Casiquiare Canal functions as a corridor allowing species distribution and gene flow between the two major aforementioned drainage systems. Secondly, the effectiveness of its function as a dispersal route is greatly influenced by its unique physicochemical and ecological nature on the one hand and the physiological and ecological tolerance of different fish species to the water chemistries of the two river systems on the other. As such, *Cichla* is possibly the best described example of a Neotropical freshwater genus whose biogeographic distribution between the Amazon and Orinoco drainages are a clear indication of the physical properties of the Casiquiare Canal (Willis et al., 2010; Winemiller and Willis, 2011). The topic of this study, *Pterophyllum,* is another fish genus that is potentially influenced by the re-established linkages between the Amazon and Orinoco drainages via the Casiquiare Canal. Of its three species both *Pt. scalare* and *Pt. leopoldi* are confined to the Amazon drainage whilst *Pt. altum* can be found in the Orinoco-, Casiquiare- and possibly the upper Rio Negro of the Amazon drainage basins (Albert and Reis, 2011). Accordingly, as a unique biogeographic corridor, the Casiquiare Canal can shed light on how ecology and historical changes in geography have influenced the distribution of Neotropical aquatic biota.

2.10 MOLECULAR MARKERS

In the initial investigations into both inter- and intrafamilial relationships of the family Cichlidae, the majority of phylogenetic studies incorporated morphological data sets to infer their phylogenetic hypotheses. However, more recently, biologists rely on the use of DNA sequence data of mitochondrial and nuclear genes which are subjected to evolutionary change through the accumulation of mutations (i.e. point mutations, insertions, deletions). In addition, the use of fossils for the calibration of time-based phylogenies has become standard practice for evolutionary biologists. Thus, 'total evidence' analyses enable the resolution of phylogenetic relationships with ever increasing accuracy and higher resolution. The most recent of such 'total evidence' analyses on Cichlidae includes the work of (López-Fernández et al., 2013, 2010), McMahan et al. (2013) and Friedman et al. (2013).

2.10.1 THE MITOCHONDRIAL GENOME

Genes encoded in the mitochondrial genome are mainly involved in the conversion of substrates to yield cellular energy in the form of adenosine triphosphate (ATP); the main energy currency of all eukaryotic cells. Hence, their protein products are of utmost importance to the metabolism and ultimate survival of eukaryotic organisms. Although the organisation of the mitochondrial genome differs among vertebrates, its genetic content is largely conserved and can be divided into three categories based on the function of their resulting gene products (**Figure 2.22**). These include functional RNA genes as well as protein-coding genes. The genes that encode functional RNA include the 16S rRNA and 12S rRNA as well as several tRNA-encoding genes. The protein-coding genes include the NADH dehydrogenase (ND) subunits, Cytochrome *c* Oxidase (COX) subunits, ATP synthase (ATP) subunits and Cyt *b*. These genes are commonly used in phylogenetic studies to aid in the elucidation of higher-level relationships. In addition, the mitochondrial genome also contains non-coding control region or "d-loop" which are highly variable and are thus frequently used especially in population studies (Pereira, 2000).

The mitochondrial genome possesses a number of unique characteristics. First and foremost, mitochondrial DNA (mtDNA) is contained within the mitochondrion, an organelle confined to the cytosol of eukaryotic cells. As mentioned before, mtDNA is haploid and consists of only one chromosome which is a small circular DNA molecule with size ranging between 15.7 – 19.5 kb (Hartl and Clark, 2007). It is double-stranded and consists of an outer (H) and inner (L) strand referred to as heavy and light strands respectively, according to their varying densities. The mitochondrial genome contains very few noncoding intergenic regions with some nucleotide overlaps between genes encoded on opposite strands (Pereira, 2000). Hence, the mitochondrial genome displays compact gene packing. Furthermore, mtDNA does not undergo recombination which means that its genetic variability is only the consequence of the accumulation of mutations.

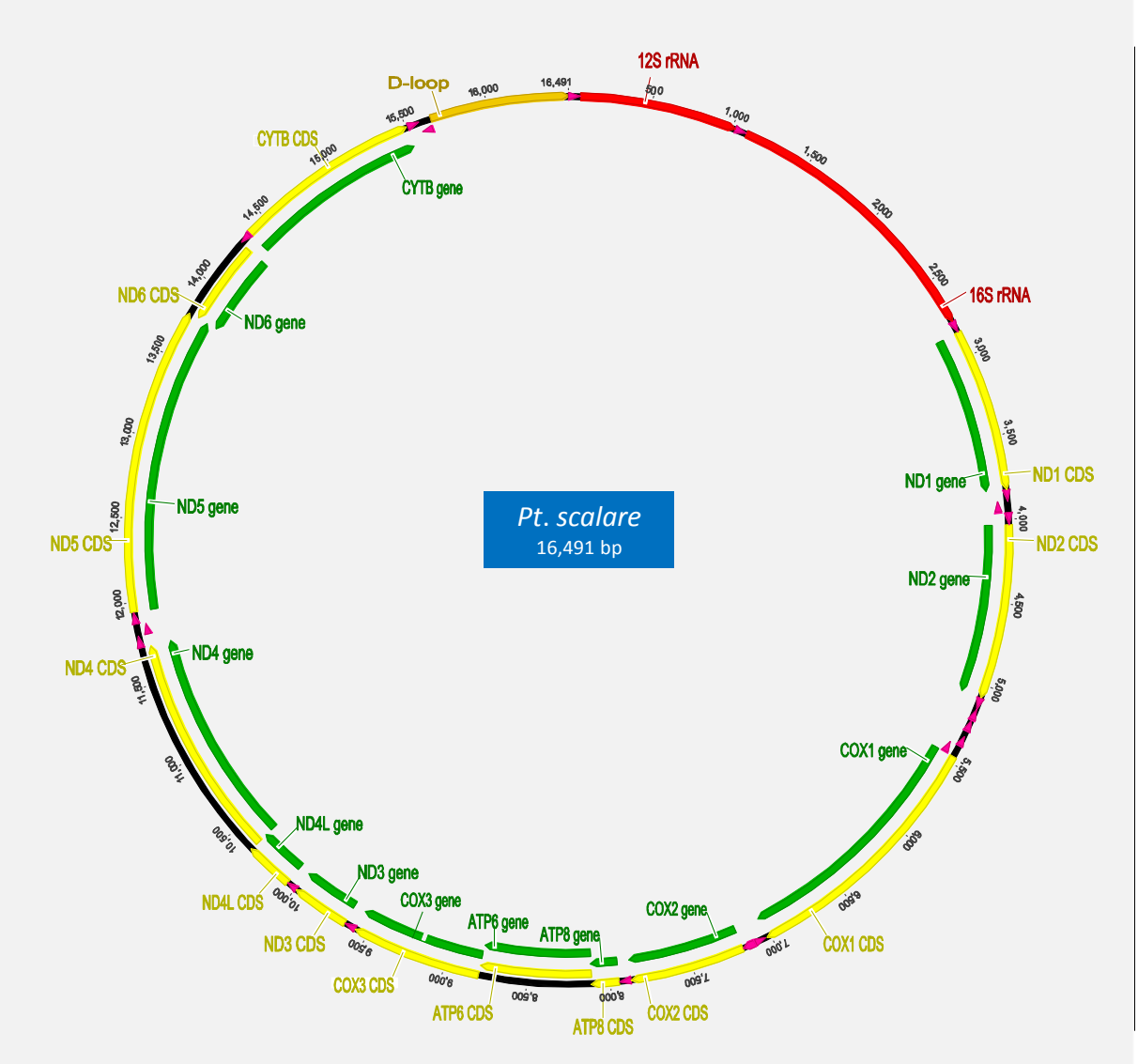


FIGURE 2.22 The mitochondrial genome of the genus *Pterophyllum*. Illustrated are the locations of all the coding regions as can be expected for the mitochondrial genome of the genus *Pterophyllum*, more specifically, the species *Pt. scalare*. The genome size is 16,491 bp. The protein-coding genes are annotated in green, ribosomal rRNA are annotated in red, transfer rRNA are marked with purple arrow heads and the D-loop is annotated in gold. RefSeq: NC_026535.1 and GenBank accession: KP231206.1 (Hao et al., 2015).

Accordingly, this corresponds to its fast mutation rate which is thought to be due to either a high rate of nucleotide disincorporation or the low efficiency of the mitochondrial DNA polymerase (Hartl and Clark, 2007). The latter lacks proofreading ability which results in ineffective DNA repair during replication. Consequently, these errors, obviously subject to selection, are transcribed and translated which contributes to a rapid rate of molecular sequence evolution. Furthermore, the mitochondrial genetic code is very similar to the universal genetic code except for the translation of the four codons: AGA, AGG, AUA and UGA (**Table 2.1**). Additionally, there are multiple copies (100 – 1,000) of the mitochondrial genome in each eukaryotic cell. The latter is particularly advantageous during DNA extraction since a percentage of the starting genetic material is lost due to enzymatic digestion. Also of interest is the fact that mtDNA is predominantly maternally inherited, thus passed on from the mother to her offspring. Therefore, because of these unique characteristics, mtDNA is deemed suitable for use

in phylogenetic studies to effectively resolve evolutionary relationships of especially those linages with relatively short divergence times (shallow nodes) and between closely related species (Kocher and Stepien, 1997; Simons et al., 1994).

TABLE 2.2 A summary of the key translational differences between the mitochondrial and universal genetic codes. Summarised are the translation results of the four triple codons AGA, AGG, AUA and UGA according to both the mitochondrial and universal genetic codes.

	Genetic code:	
Triplet codon	Mitochondrial	Universal
AGA	Termination	Arginine
AGG	Termination	Arginine
AUA	Methionine	Isoleucine
UGA	Tryptophan	Termination

The use of mitochondrial DNA sequences to address the higher-level phylogenetic relationships of the family Cichlidae, dates back to the late 1990's. During this time, mitochondrial genes were praised for their 'resolving power' as phylogenetic markers, especially the mitochondrial Cyt *b* gene. According to Farias et al. (2001) the Cyt *b* gene either alone or in combination with other data sets is able to infer phylogenetic relationships at various taxonomic levels in the family Cichlidae. However, due to its relatively fast substitution rate it is less effective in resolving deeper phylogenetic relationships than resolving the phylogenetic relationships of more recently diverged taxa. Nonetheless, the Cyt *b* gene is considered the most reliable and therefore most frequently utilised mitochondrial gene in evolutionary and phylogenetic studies pertaining to the Cichlidae (Farias et al., 2001).

Another frequently used mitochondrial gene in cichlid phylogeny is the large ribosomal subunit 16S rRNA gene. Due to its functional importance it is highly conserved across all cichlid genomes which is in accordance with its slower substitution rate compared to Cyt *b* (Simons et al., 1994). Hence, in spite of homoplasy, the 16S rRNA gene is considered a reliable phylogenetic marker to resolve deeper phylogenetic nodes and thus phylogenetic relationships between more distantly related cichlid taxa (Farias et al., 1999; Kocher and Stepien, 1997). Even though the protein-coding ND4 gene is one of the fastest evolving mitochondrial genes and considered a good phylogenetic marker (Zardoya and Meyer, 1996), it has been less frequently utilised to elucidate the phylogenetic relationships of cichlid fish. Another conservative mitochondrial gene is the cytochrome *c* oxidase subunit 1 (COI) gene (Zardoya and Meyer, 1996). In context of cichlid phylogeny, it appears that these four genes form the standard set of mitochondrial molecular markers used to infer and elucidate cichlid evolutionary relationships at various taxonomic levels.

2.10.2 THE NUCLEAR GENOME

By contrast, the nuclear genome is of necessity much more complex in its highly organised structure, various levels of gene organization, multiple genome functions and genetic variability. The nuclear genome of eukaryotic cells is contained within linear chromosomes which are organised into a dense network of fibres known as chromatin through the association with a number of histone proteins. The full complement of chromosomes present in a eukaryotic cell is determined by the cell's ploidy and chromosome number, both which can vary among eukaryotes. Animal somatic cells are diploid and hence contain two copies of each chromosome. In contrast to inheritance of the mitochondrial genome, the nuclear genome is bi-parentally inherited and each locus is represented by two alleles, one inherited from the mother and the other from the father. Furthermore, nuclear DNA can be classified as either non-repetitive or repetitive DNA (**Figure 2.23**).

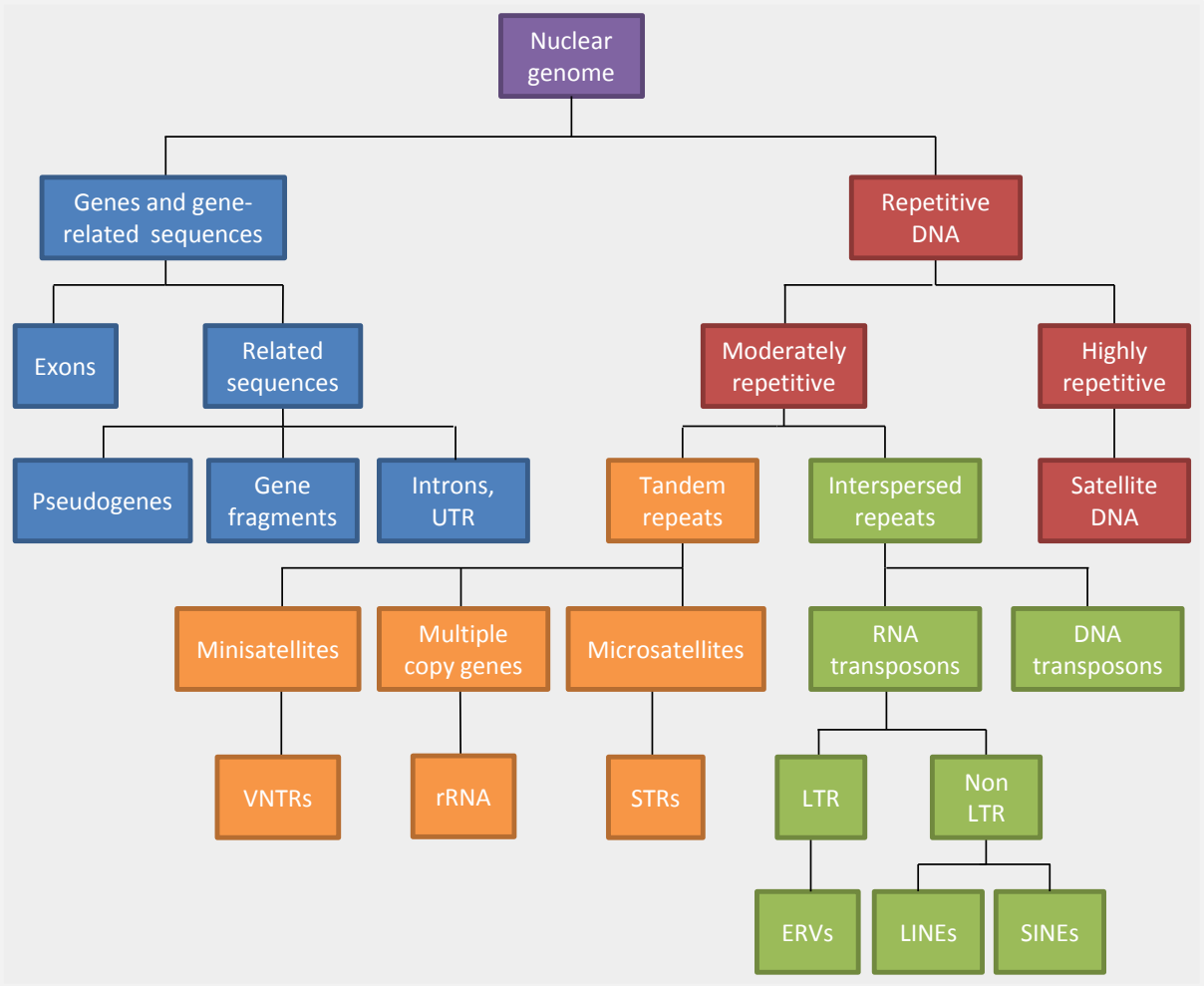


FIGURE 2.23 Nuclear genome organisation. Typical composition of a eukaryotic nuclear genome (Brown, 2007; Klug et al., 2009; Pathak and Ali, 2012).

Non-repetitive DNA refers to those DNA sequences of which only one copy exists in the genome and hence can also be called single-copy nuclear DNA. Accordingly, non-repetitive DNA sequences refer to all protein-coding DNA (exons) and their non-coding related sequences such as untranslated regions (UTRs), introns, pseudogenes and other relevant gene fragments. Protein-coding genes constitute only

a small proportion of the nuclear genome. In contrast, repetitive DNA sequences constitute much more of the nuclear genome and consequently contribute greatly to the substantial differences in genome size observed among eukaryotes (Brown, 2007). Repetitive DNA sequences or intergenic regions refer to those non-coding DNA sequences of which there are multiple copies within the nuclear genome. They can be classified into two distinct groups; highly repetitive or moderately repetitive each with various subdivisions and types (**Figure 2.23**). Characteristic of repetitive DNA is its high variability as it varies considerably in its number of basic repeat units, overall length and arrangement (Brown, 2007; Klug et al., 2009). It can either occur in tandem as an array of successive repeat units or can be interspersed throughout the genome (**Figure 2.24**). Non-coding repetitive sequences are thus dynamic DNA elements that can alter a host's genome through rearrangements, shuffling of genes and by modulating patterns of expression (Pathak and Ali, 2012). In addition to the process of natural selection, which can select against deleterious mutations, very effective DNA repair systems also contribute significantly to lower the error rate at which mismatches occur. As such, coding DNA sequences are highly conserved throughout eukaryotic lineages. Consequently, nuclear genes and functional RNA molecules have been frequently used to elucidate cichlid evolution to resolve the phylogenetic relationships of cichlids at various taxonomic levels (Fisher-Reid and Wiens, 2011).

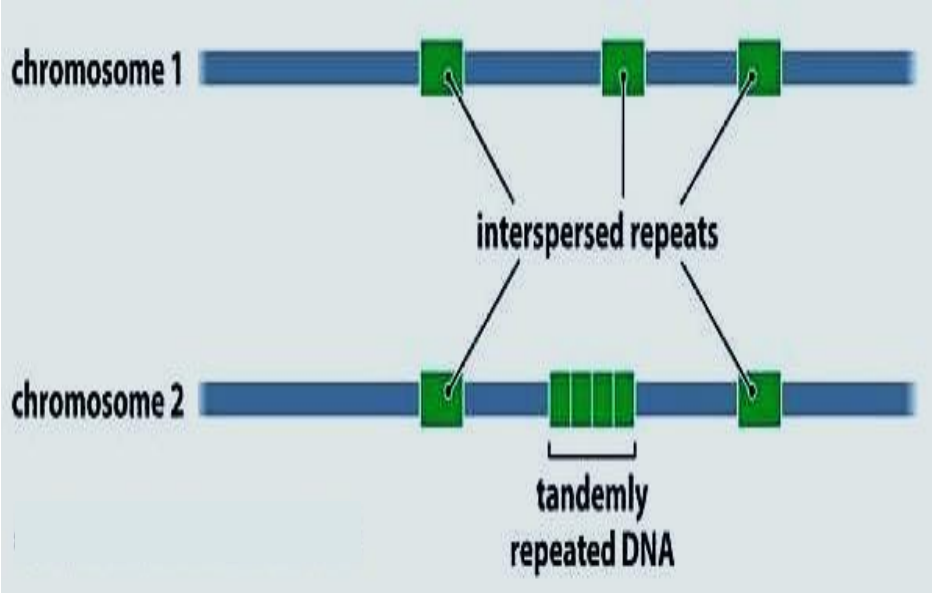


FIGURE 2.24 Moderately repetitive DNA. The typical arrangement of interspersed repeats and tandemly repeated DNA within the nuclear genome (Brown, 2007).

Amongst the vast majority of nuclear genes utilised to date, Recombinase Activator Gene (RAG) 2 and histone H3 have been frequently used (Lopez-Fernandez et al., 2005; López-Fernández et al., 2010; McMahan et al., 2013; Musilová et al., 2008; Smith et al., 2008). RAG2, in combination with RAG1, catalyses the V(D)J recombination of immunoglobulin genes (Brown, 2007; Klug et al., 2009). In contrast histone H3, in combination with histone H4, forms one of the two subunits of a nucleosome which is responsible for the packing of nuclear DNA into chromosomes during mitosis (Brown, 2007; Klug et al., 2009). Due to these genes being so conserved, they have been particularly useful to resolve deeper phylogenetic relationships of distantly related taxa.

In spite of a lack in understanding of the functions of introns (Brown, 2007), they have been frequently utilised of which the best example is the first intron of the ribosomal S7 gene (López-Fernández et al., 2010; Smith et al., 2008). Furthermore, microsatellites have gained increasing attention due to their repetitive nature and because they are highly polymorphic as a result of replication slippage and therefore highly variable (Pathak and Ali, 2012). Hence, these DNA elements are especially applicable in population genetics in particularly to resolve phylogenetic relationships between closely related taxa (Kocher & Stepien 1997). Typically, the microsatellite flanking region Tmo-M27 together with the single copy nuclear DNA locus Tmo-4C4 have also been frequently used to elucidate higher-level evolutionary relationships of cichlids (Farias et al., 2000; Karl and Streelman, 1997; López-Fernández et al., 2005; Smith et al., 2008; Streelman et al., 1998; Zardoya et al., 1996a).

2.11 THE HYBRIDISATION SCENARIO

Although older perceptions, that hybridisation results in infertile offspring in animals, are still commonly adhere to, recent evidence shows that hybridisation can lead to fertile offspring if the parent taxa are not too distantly related. For example, many Lake Malawi cichlids can give rise to fertile offspring (Joyce et al., 2011). Recent hybridisation in the cichlids from Lake Victoria has recently been shown by RAD-sequencing (Keller et al., 2013). In recent years many different phenotypic variants of *Pterophyllum* have been observed, especially within the upper Rio Negro and in regions of Santa Isabel, Lago Manacapuru and rivers in Peru. With regard to the geological history of both the Amazon and Orinoco drainages together with their presumed recent reconnection via the Casiquiare Canal, the question is whether these phenotypic variants of *Pterophyllum* might represent hybrids between *Pt. altum* and *Pt. scalare*. Hybridisation has implications for the processes of mitosis (**Figure 2.25**) and meiosis (**Figure 2.26**) in diploid organisms and is a form of recombination.

During interphase (**Figure 2.25**), the phase that precedes mitosis, the nuclear genome is replicated (Brown, 2007). At the onset of mitosis, chromosomes condense and the diploid nucleus undergoes cell division. At late prophase, the newly replicated daughter chromosomes remain attached by their centromere until metaphase when the nuclear membrane disintegrates and the chromosomes are aligned in the center of the somatic cell via microtubules. During anaphase, the microtubules draw the daughter chromosomes to either end of the cell at which time the cytoplasm starts to divide. Telophase marks the phase during which the nuclear membranes re-form for each of the identical diploid daughter cells which contain the same genes as the diploid parent cell.

In contrast, meiosis (**Figure 2.26**) only takes place in the reproductive cells and involves two consecutive nuclear divisions to generated haploid gametes. The first nuclear division involves the alignment of homologous chromosomes to form a bivalent. The chromatids of the homologous chromosomes can undergo recombination or crossing-over between bivalents which involves the physical exchange of segments of DNA. The first nuclear division thus results in two daughter cells

each with two copies of each chromosome. Meiosis then proceeds with the second nuclear division occurring in exactly the same manner as mitosis except that it results in four haploid gametes, thus, reproductive cells with only one copy of each chromosome.

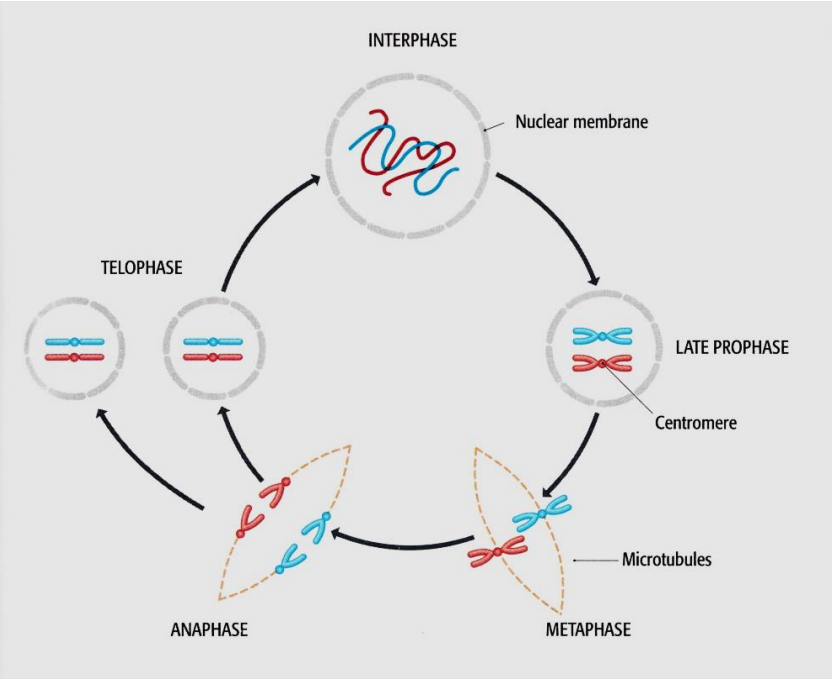


FIGURE 2.25 Mitosis. The process of mitosis involves four distinct and consecutive phases namely late prophase, metaphase, anaphase and telophase as it occurs within animal cells (Brown, 2007).

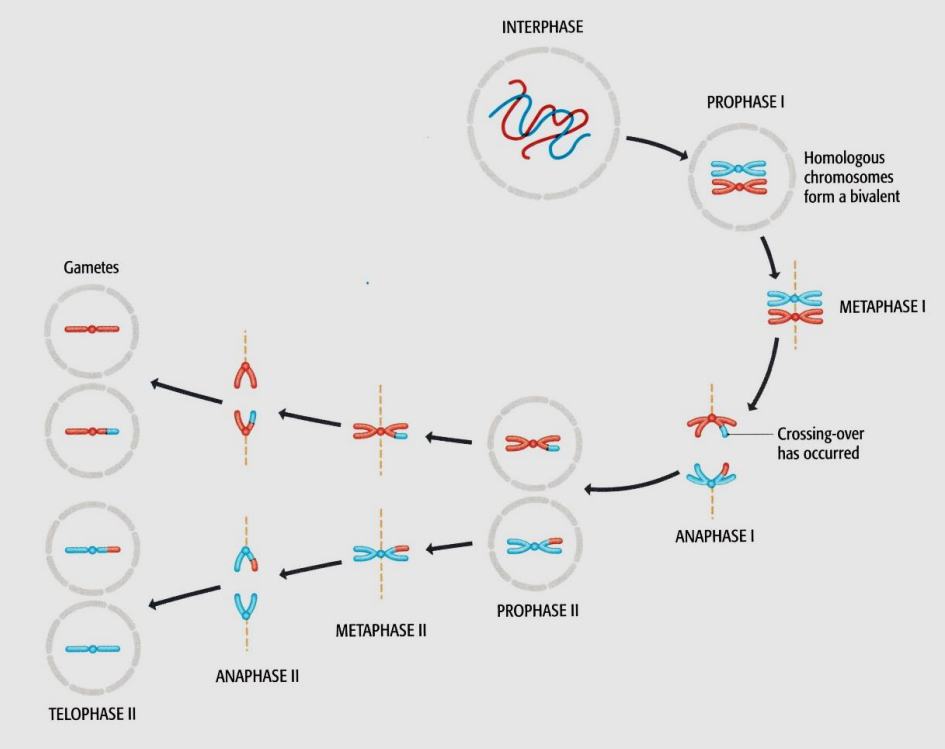


FIGURE 2.26 Meiosis. The process of meiosis, as it occurs within animal cells, involves two nuclear divisions (Brown, 2007). The first involves the division of the chromosome number such that it gives rise to two daughter cells each containing half of the nucDNA complement. Furthermore, the first nuclear division is also characterised by a cross-over event (recombination) during metaphase 1. The second nuclear division corresponds to the course of mitosis and results in four haploid gametes.

Hybridisation refers to the biological process during which individuals of two species mate. In a hypothetical dihybrid cross between *Pt. altum* and *Pt. scalare* the alleles of gene S (S and s) and gene R (R and r) may be considered (**Figure 2.27**). When considering the cross between a male *Pt. altum* (rrSS) and female *Pt. scalare* (RRss), the genotype of the first (F1) generation will be heterozygous for both genes (RrSs) and will contain the mitochondrian of the female i.e. *Pt. scalare*.

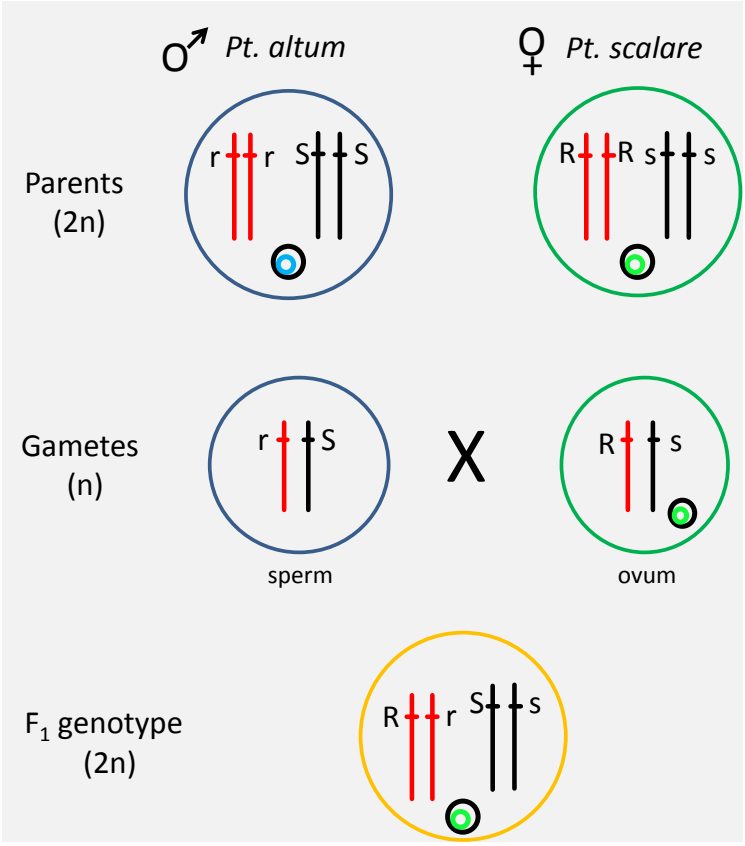


FIGURE 2.27 Hybridisation. A parental cross between two species of the genus *Pterophyllum* and their resulting hybrid offspring (F_1) containing the mitochondrion of the mother which is maternally inherited.

In the reverse cross between a female *Pt. altum* and male *Pt. scalare* the offspring would then contain the mitochondrion of *Pt. altum*. In turn, when two hybrid F1-generation fish produce further offspring, a number of different offspring genotype combinations could be produced in the F2 and subsequent generations (**Figure 2.28**). If further individuals of either parent species are not introduced again, this could lead to the formation of hybrid populations with fixed ratios of alleles regulated according to the Hardy-Weinberg equilibrium. Further processes such as replication and recombination could create additional genetic variability in the nuclear genome which may further influence allele frequencies within a population. In small populations, the allele fixation rate is rapid which means that although such populations are hybrids they may become divergent in a short space of time. Hence, should hybridisation occur within different *Pterophyllum* populations, the numerous offspring genotype combinations could potentially result in a large variety of phenotypes across hybrid populations. Thus the large variety in phenotypes observed in *Pterophyllum* populations in the upper and lower Rio

Negro may possibly be the result of such hybridization events followed by allele fixation in isolated populations. Members of the genus *Pterophyllum* are known to frequent quiet backwater tributaries in smaller populations, and not occur in big shoals in the mainstreams of the big rivers such as the Rio Negro (Linke, personal communication). However, these populations are not entirely isolated because the main stream of the Rio Negro does connect the side tributaries allowing individuals to move from one tributary to the next, particularly during the annual flood that is found in these river systems.

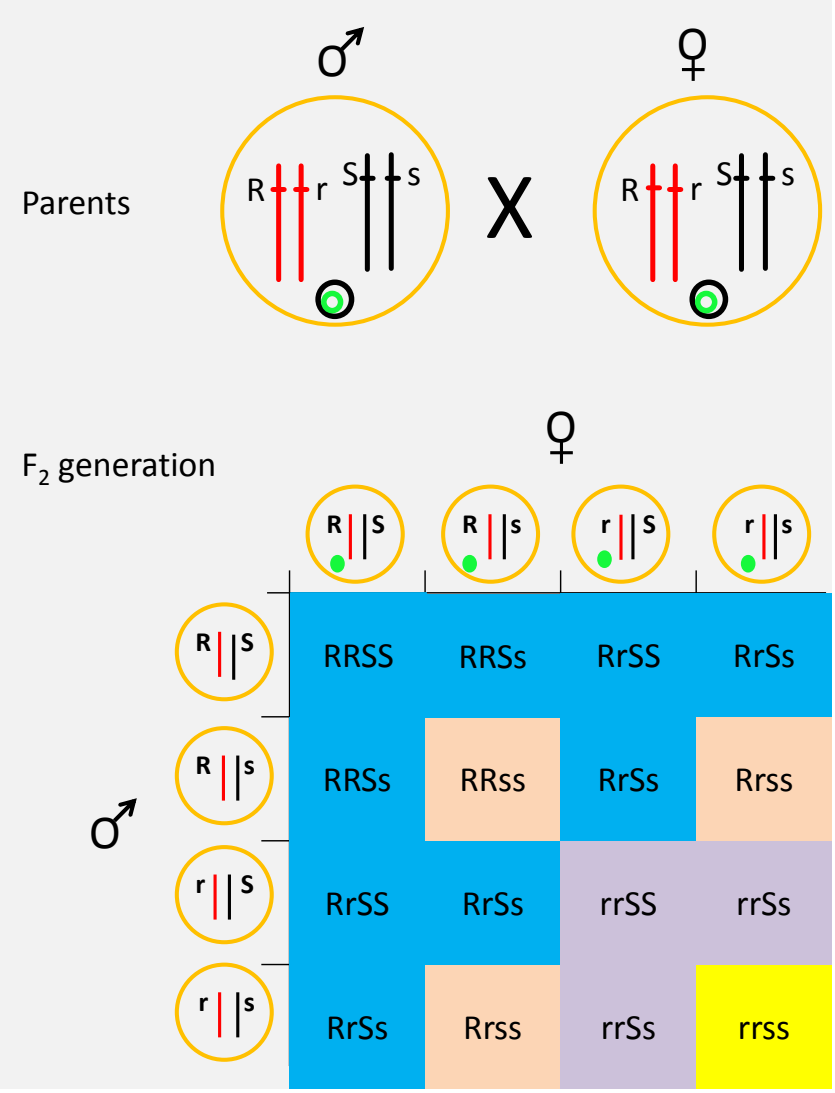


FIGURE 2.28 Various F_2-genotype combinations. A simplified representation of the allele combinations resulting from crosses between two F_1 individuals resulting in 16 genotype combinations when two unlinked genes are considered.

In the context of this study, hybrid populations of *Pt. scalare* and *Pt. altum* may therefore be indicated by different combinations of the haploid mitochondrial genes of the one species with nuclear genes of the other species, although allele fixation of the nuclear genes to the one or the other parent species may obscure such hybridization events. In order to establish unequivocally whether hybridization has occurred, whole genome scans, such as can be performed using the RAD Sequencing (Davey and Blaxter, 2010) would have to be applied.

2.12 MOLECULAR PHYLOGENETICS

Molecular phylogenetics is the science of estimating the evolutionary relationships between different species or genes through the use of both molecular techniques and computational methods (Dowell, 2008). The evolutionary information of all living organisms is organized and contained within the complex macromolecules such as nucleic acids (DNA/RNA) and proteins which retain a record of an organism's evolutionary history. Through the field of molecular biology, genetic information related to the structure and function of these macromolecules and how they change over time, through the accumulation of mutations and recombination can be used to determine the evolutionary relationships among species or genes (Dowell, 2008; Vandamme, 2009). One advantage of the molecular approach in determining phylogenetic relationships over the more classical approaches, such as those based on morphological characters or life cycle traits, is that the differences are readily quantifiable. Molecular data can thus provide valuable insights that can assist species identification and higher-level phylogenetic analyses (Vandamme, 2009). Therefore, by constructing a comprehensive and well-resolved phylogeny, one is providing the necessary scaffold upon which the evolutionary history of lineages can be traced. This in turn enables the use of molecular data to not only estimate temporal divergence and understand the biogeographic history of lineages, but to also resolve taxonomic confusion. To this end, evolutionary processes responsible for biological diversity can be better understood (Hulsey et al., 2011; Karl and Streelman, 1997; McMahan et al., 2013; Roe et al., 1997; Sparks, 2004; Streelman et al., 1998)

2.12.1 PHYLOGENETIC TREES

A phylogeny is, in true nature, a complicated estimate of the evolutionary development and history of species which are graphically represented by a phylogenetic tree. There are a few characteristic features that form the basis of any given phylogenetic tree. The taxa whose phylogenetic relatedness is being inferred are referred to as the external or terminal nodes. Sometimes, these nodes are also called operational taxonomic units (OTUs). Likewise, a tree consists of many internal nodes which can individually be considered as the progenitor of all the nodes that arose from it. Accordingly, all nodes are connected via branches. The unique branching pattern (order of the nodes) is referred to as the topology of the tree. If a group of taxa are all descendants of the same common ancestor (internal node), the group is said to be monophyletic and hence form a clade. In contrast, a paraphyletic group consists of the majority of taxa from one clade and one or two taxa from a neighbouring clade. Therefore, these taxa are not derived from the same common ancestor. In addition, the taxa form a polyphyletic group are distantly related as they represent different clades, but are associated because they share certain phenotypic traits albeit not inherited from the same common ancestor.

Phylogenetic trees can either be rooted or unrooted. In context of a rooted tree, the root can consist of either a single or a group of taxa known as the outgroup species, which serves as reference to which all

other taxa, the ingroup, can be compared. To this end, the outgroup taxa are generally chosen to be the most distantly related of all the OTUs. Thereby, the outgroup serves the purpose to indicate the direction of evolution in a tree. Increased taxon sampling of outgroups generally improves ingroup resolution (Vandamme, 2009). However, when there is no suitable outgroup available but the rate of evolution across different lineages is similar, a root can still be assigned through the process of midpoint rooting (Vandamme, 2009). In the case of an unrooted tree, there is no common ancestor and subsequently the taxa are positioned relative to one another. In order to generate a phylogenetic tree for a given set of aligned nucleotide sequences, computational methods are applied to calculate the branch lengths which describes the phylogenetic relationships of the nucleotide sequences and determine the best tree topology (Dowell, 2008).

2.12.2 COMPUTATIONAL METHODS OF PHYLOGENETIC ANALYSES

According to Hartl & Clark (2007), "*The problem of inferring the correct branching topology for a tree that relates a set of organisms is a challenge in part because of the enormous number of possible bifurcating trees".* In a rooted bifurcating tree, each internal node gives rise to only two descendants. *"If there are n species to be placed, there are* $(2n\text{-}3)!/2^{n-2}(n\text{-}2)!$ *rooted trees that describe possible ancestral histories. For many data sets of 30 or more species, the number of possible trees is so enormous that it is not possible to examine all topologies and assess the fit of the data to each tree, even with the very fastest computers. Fortunately, the trees are not all independent of one another, and the key to many of the algorithms that try to find the best-fitting tree is to eliminate whole classes of trees based on the observed data*.".

Methods for inferring phylogenetic relationships are either pairwise-distance or discrete-character methods (Swofford and Sullivan, 2009). The former utilises the information contained within a matrix of pairwise distances where each distance is an estimation of the amount of divergence between two taxa since their most recent common ancestor (MRCA) (Swofford and Sullivan, 2009). In contrast, discrete-character methods utilise the information from an alignment of sequences, albeit DNA or protein, where each taxon is defined by its unique set of character states. The most commonly discrete-character methods used for phylogenetic inference are parsimony and maximum likelihood (ML). These methods rely on optimality criteria that explicitly define unique objective functions by which they are able to score the topology of any given phylogenetic tree. Thereby, phylogenetic trees can be ranked and the optimal tree topology identified (Swofford and Sullivan, 2009).

2.12.2.1 Parsimony analysis

A very commonly method used for phylogenetic tree reconstruction is the parsimony algorithm which can be implemented in Phylogenetic Analysis using Parsimony* (PAUP*) (v 4.0b10) (Swofford, 2003). PAUP* is a computational phylogenetic program. Parsimony is an optimality criterion under which it is possible to infer evolutionary relationships from a set of aligned sequences

by identifying the tree or collection of parsimonious trees with the smallest possible number of evolutionary changes that can explain the data (Hartl and Clark, 2007; Swofford and Sullivan, 2009). In other words, parsimony is based on the heuristic search for the shortest trees that can account for their evolution by using an optimality criterion. The result of a parsimony analysis is contained within a rooted binary changes tree. The branch lengths of a changes tree represent the number of mutational events that can best describe the evolution of a given taxon in relation to the root of the tree. As such, the parsimony-criterion assumes that the shortest possible tree that can explain the data is considered the most parsimonious tree. However, there is no guarantee that the most parsimonious tree is necessarily the best or statistically most correct tree (Hartl and Clark, 2007). As there may be many equally parsimonious trees as a result, a consensus method and bootstrap analysis are implemented to assist in identifying the best possible tree to describe the data at hand.

There are four different consensus methods available in PAUP*; Strict, Semistrict, Adams and Majority-rule. Consensus methods involve identifying common elements between two or more equally parsimonious trees that is then subsequently summarised into a single tree; hence a consensus tree (Margush, 1981). The simplest and probably most commonly used consensus method is strict consensus which involves the computation of a strict consensus tree based on the frequency of occurrence of clusters in a set of trees. Therefore, the strict consensus tree contains *only* those clades or clusters that are unambiguously supported in *all* of the equally parsimonious trees (Margush, 1981). All other remaining clades or clusters are consequently represented as unresolved polytomies in the strict consensus tree. In the context of a binary tree, a polytomy represents a group of taxa whose phylogenetic relationships between one another are unclear and therefore unspecified. This is graphically represented as a node with many descendants instead of only two as per the definition of a binary tree. In such a situation it is said that the most recent common ancestor (internal node) of a polytomy has ‘collapsed’.

When constructing a phylogenetic tree, it is important to assess the statistical confidence in the nodes of the different clusters or groupings. To this end, the bootstrap method can explore the robustness (relative stability) of groups within a phylogenetic tree as it is based on a heuristic approach. The bootstrap method is a sampling technique that involves the repeated, random resampling of data of an original data set to estimate clade support and hence the statistical confidence of a tree (Felsenstein, 1985). Initially, the original data set is bootstrapped during which a subset of the original data is randomly drawn with replacement to form a bootstrap replicate or new set of data (Felsenstein, 1985). Subsequently, a tree is constructed from this new data in which each node of the new tree is compared to that of the original tree to see if the new tree has the same cluster of sequences (Felsenstein, 1985). This process of resampling data, computing trees and comparing clades, is routinely repeated 1000 times for it to be statistically robust. Once this process is complete a majority rule consensus tree is computed in which bootstrap values are given next to each node. These bootstrap values indicate the percentages that a particular cluster was present among all the resampled trees. For example, if the

bootstrap value of a node is 90, it means that the particular cluster was present in 90% of all the resampled trees. If the bootstrap value is greater than 75%, it is accepted to be statistically well supported which means that it can be accepted, with confidence, that a given cluster belongs together and hence is resolved (Hartl and Clark, 2007).

2.12.2.2 Maximum likelihood analysis

By definition, likelihood refers to the condition or state of something being probable. Accordingly, from an evolutionary point of view, by applying a likelihood analysis to a matrix of aligned nucleotide (DNA or RNA) sequences, provides a statistical manner by which to assess its probability and hence obtain an adequate explanation of the data at hand (Felsenstein, 1981). The uniqueness of any given data set can be attributed to the pattern of gene mutations observed. Hence, in order to explain the data, a nucleotide substitution model is implemented under which many hypotheses, about the data, are formulated. These hypotheses relate to the many probable evolutionary trees, each defined by its unique topology and branch lengths, in addition to the set of model parameters (Felsenstein, 1981). Therefore, the aim is to construct a phylogenetic tree i.e. formulate a hypothesis, that can best explain the evolutionary relationships of the aligned sequences with the highest likelihood.

As a model-based method, maximum likelihood achieves this through its likelihood function which estimates values for the model parameters such that it maximizes the likelihood of a hypothesis explaining the observed data. As such, the maximum likelihood method is used to infer phylogenetic relationships based on its principle "*to determine the tree topology, branch lengths, and parameters of the evolutionary model that maximize the probability of observing the sequences at hand*" (Schmidt, Heiko and von Haeseler, 2009). The model parameters and branch lengths are computed for each tree. The tree topology with the highest likelihood is then considered the best to explain the data at hand.

One of the programs used to perform phylogenetic analyses under maximum likelihood is called Randomized Axelerated Maximum Likelihood or pronounced as abbreviated; RAxML-HPC2 available on XSEDE version 8.0.24 (Stamatakis, 2014) available on The CIPRES Science Gateway V. 3.3 (Miller et al., 2010). The latter implements a fast maximum likelihood tree search algorithm that returns only the most likely tree with bootstrap support (Stamatakis, 2014). Branch lengths represent the probability of a nucleotide transitioning from one character state to another. Hence, short branches can be interpreted as a low probability for a nucleotide, in a given sequence, to change thus, a high probability to continue to exist in its given state. The opposite is true for long branches.

2.12.2.3 Model selection: how to select a model

Statistical selection of the best-fit model of nucleotide substitution can be performed by jModelTest 2.1.6 v20140903 (Darriba et al., 2012; Guidon and Gascuel, 2003) which is run on XSEDE available on The CIPRES Science Gateway v3.3 (Miller et al., 2010). It implements five different model

selection strategies: hierarchical and dynamical likelihood ratio tests (hLRT and dLRT), Akaike and Bayesian information criteria (AIC and BIC), and a decision theory method (DT) (Darriba et al., 2012; Guidon and Gascuel, 2003). Based on these selection strategies jModelTest is able to statistically identify the best-fit substitution model that can best describe the different probabilities of base frequency changes of a given data set of aligned nucleotide sequences.

2.12.2.4 Bayesian Evolutionary Analysis by Sampling Trees (BEAST)

BEAST is a software program used to perform a likelihood analysis of aligned nucleotide sequence matrices using the Bayesian optimality criterion to generate time based phylogenies by implementing the Monte-Carlo Markov Chain (MCMC) algorithm (Drummond et al., 2007) as typically used in v1.8.2 (Drummond and Rambaut, 2007; Suchard and Rambaut, 2009) run on XSEDE available on The CIPRES Science Gateway v3.3 (Miller et al., 2010). Due to the use of the Bayesian algorithm, BEAST is able the combine the prior knowledge (i.e. the prior probability distribution) and the likelihood distribution of a given multiple sequence alignment to estimate the posterior probability distribution that can best describe the data at hand (Drummond and Rambaut, 2009). This is initiated through the software program BEAUti. It requires a Nexus file, containing the multiple sequence alignment, as input. By means of BEAUti, prior knowledge of the data can be captured by setting the evolutionary model (nucleotide substitution model, site heterogeneity model, molecular clock model, tree model and tree prior) and the informative priors for their respective parameters. Furthermore, the options for the likelihood analysis using the MCMC algorithm such as the length of the chain and the sampling frequency of parameter values to be logged, is also set. The subsequent generated XML file, containing the data, evolutionary parameters and analysis settings such as chain length and frequency of logged parameters, is then used as an input file to run BEAST.

In order to estimate the posterior probability distribution of the set of evolutionary parameters, BEAST implements the stochastic Metropolis-Hastings MCMC algorithm to "*average over tree space, so that each tree is weighted proportionally to its posterior probability"* (Drummond et al., 2007). The number of samples in a trace is dependent on the specified length of the MCMC chain and the sampling frequency of parameter values. For example, if the chain length consists of 200,000 steps and parameter values are logged every 100 steps, then there will be 2,000 samples in total. The output of a BEAST analysis is contained within two separate text documents; a log and tree file. The log file contains the traces i.e. sampled probability values for the posterior and continuous parameters as per logged state. The tree file contains all the probable trees generated per logged state. Additional utility programs are included within the BEAST package that can assist in the processing of the data contained within the log and tree files. One such program is LogCombiner which is able to combine the log and tree files from independent MCMC analyses into a single log and tree file respectively. These BEAST output files are each explored by different programs.

The concatenated log file is analyzed using the program Tracer. Unfortunatly Tracer is not included within the BEAST package and therefore must be additionally obtained. Nonetheless, Tracer enables a user to graphically interpret the posterior distributions of each evolutionary parameter and provides diagnostic information (Drummond and Rambaut, 2009). Of importance to note is the ESS value for each trace, in particularly of the posterior trace. The ESS value represents the fraction of the total number of samples (see example above) that was independently sampled for a particular trace (Drummond and Rambaut, 2009). Should this value be low, it can be assumed that sampling for a given trace was not sufficient. In the case of the posterior trace, it may not represent the posterior distribution very well (Drummond and Rambaut, 2009). Hence, the higher the ESS value, the better the estimation of the posterior probability distribution. An ESS value greater than 200 is widely accepted as indicative of good support in the scientific community (Drummond and Rambaut, 2009). However, since the total number of samples can vary according to the MCMC chain length and sampling frequency, it is probably best to analyse the raw trace plot of each trace to see if the MCMC has converged, that is, whether sampling can be ruled to be adequate.

Conversely, the concatenated tree file first needs to be analysed by the utility program TreeAnnotator. TreeAnnotator is an application that summarises the information contained within the sampled trees (Drummond and Rambaut, 2009). In other words, it evaluates the tree space for the tree with the highest posterior probability and subsequently annotates it. The posterior probability of a tree can be interpreted as the likelihood that the tree is correct. As such, the tree with the highest posterior probability can be considered the best tree to reconstruct a phylogeny or test an evolutionary hypothesis. The information contained within the annotated tree include average node ages specified by the 95% HPD upper and lower bounds, posterior probability support (values ranging from 0.1-1.0) and the average rate of evolution on each branch (Drummond and Rambaut, 2009). This resulting annotated tree can then be visualised with the program FigTree, which must be obtained separately from the BEAST package. It can thus be concluded that the result of a BEAST analysis is the most probable, rooted, time-measured phylogenetic tree that best describes the data at hand given prior knowledge thereof.

3. A FIVE-GENE PHYLOGENY OF THE GENUS *PTEROPHYLLUM* WITHIN THE CICHLIDAE

3.1 INTRODUCTION

Of all the somewhat 1600 cichlid species of the fish family Cichlidae, the South American genus *Pterophyllum*, or commonly known as Angelfish, has attracted the attention of ichthyologists and aquarists alike because of their striking morphological features which include being laterally compressed and having extremely elongated dorsal and ventral fins. Presently, it is accepted that the genus *Pterophyllum* (Sub-family: Cichlinae; Tribe: Heroini) consists of three species; *Pt. scalare* (Schultze, 1823), *Pt. altum* (Pellegrin, 1903) and *Pt. leopoldi* (Gosse, 1963). Biogeographically, these species are widespread throughout the Amazon, Orinoco and Essequibo drainage basins. Despite the knowledge of its widespread distribution, the genus *Pterophyllum* has been underrepresented at species-level in the majority of phylogenetic studies focused on Neotropical cichlids (Concheiro Pérez et al., 2007; Farias et al., 2001, 2000, 1999; Kullander, 1998; López-Fernández et al., 2010; Říčan et al., 2013; Smith et al., 2008; Sparks, 2004; Stiassny, 1991)

To address the lack of phylogenetic information, Meliciano (2009) conducted a phylogeographic study of the genus *Pterophyllum* based solely on the mitochondrial Cyt *b* gene. In addition, her study included morphometric data obtained from an extensive taxon sampling of *Pt. scalare*, *Pt. altum* and *Pt. leopoldi*. However, the sampling locations for what she viewed as *Pt. altum* were restricted to the areas of Santa Isabel and Boa Vista in the upper Rio Negro. As such, *Pt. altum* collections from its type locality i.e. Rio Atabapo in addition to sampling locations from the Upper Rio Orinoco and Rio Casiquiare, were not included. Although limited to a gene tree and morphometric measurements, this study gave the first insights into relationships in the genus *Pterophyllum*. Furthermore, outgroup taxa from closely related Heroni genera were limited to *Symphysodon discus*. Consequently, the phylogenetic position of the genus *Pterophyllum*, within the tribe Heroini, was not addressed.

The first aim of this study was therefore to address and resolve the relationships of the three *Pterophyllum* species i.e. *Pt. altum*, *Pt. scalare* and *Pt. leopoldi* based on more sequence information. The second aim was to formulate a robust phylogenetic hypothesis for the position of the genus *Pterophyllum* relative to its immediate neighbouring genera in the tribe Heroini. The third aim was to identify sufficiently variable mitochondrial and nuclear markers to further explore intraspecies variability within the species groups. The final aim was to hypothesize dates for the temporal divergence of the three species of the genus *Pterophyllum.*

In an attempt to achieve these aims, the work by López-Fernández et al., (2010) was used as the platform for the construction of a five-gene phylogeny for the genus *Pterophyllum*. This phylogeny was based on five molecular markers; three mitochondrial (Cyt *b*, 16S rRNA and ND4) and two

nuclear (S7 and RAG2) in addition to an extensive taxon set of 154 Neotropical cichlids (López-Fernández et al., 2010). The five genes were sequenced from a number of taxa representative of the three species of the genus *Pterophyllum* and their sequences added to the sequence matrices of López-Fernández in order to generate a phylogeny. The taxa included *Pt. altum* samples from its type locality i.e. Rio Atabapo and other tributaries of the Upper Rio Orinoco as well as samples of putative *Pt. altum* from the upper Rio Negro. Furthermore, taxa of *Pt. scalare* from a number of localities in the Amazon River, Santa Isabel in the upper Rio Negro and Lago Manacapuru from the vicinity of Manaus as well as *Pt. leopoldi* were included. The character variability of the five (Cyt *b*, 16S rRNA, ND4, S7 and RAG2) sequence matrices was assessed separately in order to identify the most variable mitochondrial and nuclear marker genes. Finally, a BEAST analysis was performed on the mitochondrial Cyt *b* sequence matrix in order to assess the termporal divergence of the three *Pterophyllum* species groups within the context of the 154 Neotropical cichlid taxa.

3.2 MATERIALS AND METHODS

3.2.1 SAMPLE COLLECTION

In total, 27 tissue samples were sourced from various localities by Mr Edgar Ruiz, Dr. Helmut Wedekind and Dr. Dirk Neumann for use in this study (**Table 3.1**). Mr. Edgar Ruiz, a trained Venezuelian ichthyologist based in the U.S.A., obtained *Pterophyllum* samples from reliable fish collectors in Venezuela. These included juvenile *Pt. altum* fishes from the Rio Inírida, Rio Atabapo and Rio Ventuari. Fin clips of a number of adult *Pt. altum* fishes, imported into the U.S.A. from San Felipe (Colombia, on the opposite bank of the river to San Carlos in Venezuela) on the upper Rio Negro, were also included (**Figure 3.1**). Dr. Helmut Wedekind, Head of Fisheries, Department of Agriculture, Bavaria, Germany, and a keen aquarist, supplied *Pterophyllum* samples from his live collection. These samples were sourced from reliable sources in the aquarium trade in Germany and included fin clipped samples of *Pt. altum*, *Pt. scalare* and *Pt. leopoldi*. *Pterophyllum scalare* samples from domestic stock, which in all likelihood originated from Manaus, were indicated as *Pt. scalare* Manaus. Dr. Dirk Neumann of the Bavarian State Collection, Ichthyology Section, supplied samples of *Pt. scalare* and *Pt. leopoldi* from Aquarium Glaser, a leading tropical fish importer in Germany. The collection of 27 tissue samples was representative of the three species of the fish genus *Pterophyllum* namely *Pt. altum* (10), *Pt. scalare* (15) and *Pt. leopoldi* (2). According to procedure, a small tissue sample from the caudal/tail fin of each fish was collected and preserved in 99% ethanol. However, samples from Dr. Dirk Neumann were supplied as lyophilised DNA samples.

3.2.2 DNA EXTRACTION

Initially, DNA was extracted using a slightly modified Phenol-Chloroform-Isoamylalcohol protocol (Dorn et al., 2011). In short, samples that had been stored in ethanol were washed by adding 500 µL Milli-Q® water to each tube and then pulse-vortexed for 5-10 seconds. The wash step was repeated to

ensure the removal of any residual ethanol. Each sample was then placed in a 2 mL Eppendorf tube containing 600 µL digestion buffer (10 mM Tris, pH 8.0, 100 mM NaCl, 10 mM EDTA, 0.5% w/v SDS) and 20 µL of Proteinase K solution (50 µg/mL). Samples were then incubated overnight at 55°C on a heating block (AccuBlock, Labnet International, Inc.) to allow digestion of the protein in the tissue samples to release the DNA.

Following incubation, 600 µL of a solution of phenol-chloroform-isoamylalcohol (25:24:1) was added to each tube, vortexed for three seconds and centrifuged at 13,500 x *g* for 10 minutes. The resulting aqueous phase was removed and transferred to a new 2 mL Eppendorf tube. This step was then repeated by adding 600 µL of a chloroform-isoamylalcohol (25:1) solution. The subsequent aqueous phase was removed, placed into a 2 mL Eppendorf tube containing 1 mL of ice cold 99% ethanol to allow the DNA to precipitate and then centrifuged at 13,500 x *g* for 10 minutes. Each resulting pellet was washed in 500 µL 70% ethanol, stored at -20°C, and then centrifuged at 13,500 x *g* for 5 minutes. Ethanol was removed and the samples air-dried at 37°C after which each pellet was re-suspended in 50 µL TE-buffer (10 mM Tris, 1 mM EDTA, pH 8.00). The samples were then incubated at 37°C overnight to allow the pellet to dissolve. After a number of PCI-extractions had been performed it was found that the DNA yield of many of the samples was very low (**Table 3.1**). Since these tissue samples were limited, DNA extraction could not be repeated using more starting material. Therefore, all subsequent extractions were performed using the DNeasy Blood and Tissue kit (Qiagen) according to the supplier's protocol with a slightly modified elution strategy. Instead of using 100 µL of elution buffer for both the first and second elutions, 120 µl was used for the first elution and 80 µl for the second elution step (**Table 3.1**).

FIGURE 3.1 A map showing collection localities on the Upper Rio Negro. Annotated are the neighbouring cities of San Carlos (Venezuela) and San Felipe (Colombia) which are located a few kilometres drownstream from where the Rio Casiquiare conjoins with the Rio Negro. From here, the Rio Negro runs along the Colombian/Venezuelian border and crosses into Brazil in the vicinity of La Guadalupe, eventually passing the regions of Boa Vista and Santa Isabel. Image from Google Maps.

TABLE 3.1 A list of the 27 tissue samples utilised in this study. Included is a summary of the source of origin, method of DNA extraction, the resulting DNA concentrations and molecular markers analysed per sample.

Sample number	Sample code	Species names	Source	Method of extraction*	[DNA] (µg/µL)	Molecular Markers Analysed:				
						16S rRNA	Cyt *b*	ND4	S7	RAG2
	***Pt. leopoldi*:**									
1	PAA-766	*Pt. leopoldi* C	Aquarium Trade	DNeasy	-	X	X	X	X	X
2	PAA-767	*Pt. leopoldi* D	Aquarium Trade	DNeasy	-	X	X	X	X	X
	***Pt. altum* (Orinoco):**									
3	JARA 1	*Pt. altum* Rio Atabapo 1	Rio Atabapo	DNeasy	5.9	X	X	X	X	X
4	JARA 2	*Pt. altum* Rio Atabapo 2	Rio Atabapo	DNeasy	5.6	X	X	X	X	X
5	JARI 1	*Pt. altum* Rio Inirida 1	Rio Inirida	DNeasy	-0.1		X	X		X
6	JARI 2	*Pt. altum* Rio Inirida 2	Rio Inirida	DNeasy	18.3		X	X	X	X
7	JARI 5	*Pt. altum* Rio Inirida 5	Rio Inirida	PCI	60.0	X				
8	JARI 6	*Pt. altum* Rio Inirida 6	Rio Inirida	PCI	89.4	X				
9	JARV 3	*Pt. altum* Rio Ventuari 3	Rio Ventuari	PCI	9.3	X	X	X	X	X
10	JARV 4	*Pt. altum* Rio Ventuari 4	Rio Ventruari	PCI	8.0	X	X	X	X	X
	***Pt. altum* (Rio Negro):**									
11	ARN 1	*Pt. altum* Rio Negro 1	Rio Negro	DNeasy	44.2	X	X	X	X	X
12	ARN 2	*Pt. altum* Rio Negro 2	Rio Negro	DNeasy	67.5	X	X	X	X	X

Sample number	Sample code	Species names	Source	Method of extraction*	[DNA] (µg/µL)	Molecular Markers Analysed:				
						16S rRNA	Cyt *b*	ND4	S7	RAG2
	***Pt. scalare*:**									
13	PAA-711	*Pt. scalare* Manaus B	Aquarium Trade	PCI	6.5	X	X	X	X	X
14	PAA-712	*Pt. scalare* Manaus C	Aquarium Trade	PCI	10.4			X	X	X
15	PAA-755	*Pt. scalare* Manaus D	Aquarium Trade	DNeasy	-	X	X			
16	PAA-713	*Pt. scalare* Guyana A	Aquarium Trade	PCI	7.1	X	X	X	X	X
17	PAA-714	*Pt. scalare* Guyana B	Aquarium Trade	PCI	6.1	X	X	X	X	X
18	PAA-703	*Pt. scalare* Santa Isabel A	Aquarium Trade	PCI	17.2	X			X	X
19	PAA-704	*Pt. scalare* Santa Isabel B	Aquarium Trade	PCI	14.7	X		X	X	X
20	PAA-763	*Pt. scalare* Manacupuru A	Aquarium Trade	DNeasy	-	X	X	X	X	X
21	PAA-764	*Pt. scalare* Manacupuru B	Aquarium Trade	DNeasy	-	X	X	X	X	X
22	PAA-705	*Pt. scalare* Peru 1A	Aquarium Trade	PCI	15.1	X		X	X	X
23	PAA-706	*Pt. scalare* Peru 1B	Aquarium Trade	PCI	8.4	X				
24	PAA-709	*Pt. scalare* Peru 2A	Aquarium Trade	PCI	4.9		X	X	X	X
25	PAA-710	*Pt. scalare* Peru 2B	Aquarium Trade	PCI	6.4		X			
26	PAA-761	*Pt. scalare* Xingu B	Aquarium Trade	DNeasy	-	X	X	X	X	X
27	PAA-762	*Pt. scalare* Xingu C	Aquarium Trade	DNeasy	-	X	X	X	X	X

***DNeasy:** DNeasy Blood and Tissue kit**; PCI:** Phenol-Chloroform-Isoamylalcohol

Upon receiving the lyophilised DNA samples from Dirk Neumann, 20 µL Milli-Q® water was added to each tube and then incubated at 37°C for 5-10 minutes to allow the DNA to dissolve. These samples were then gently mixed using a 20 µL pipette. All DNA samples were stored at - 20°C until further analyses were performed. The DNA concentration of all the samples was measured using the NanoDrop® (Inqaba Biotec). Initialisation of the machine was done using 2 µL Milli-Q® water. For the DNA samples extracted by means of the PCI-method, 2 µL TE-buffer (elution buffer) was used as the appropriate blank measurement. Similarly, 2 µL Buffer AE (elution buffer) was used as a blank for the DNA samples extracted with the DNeasy Blood and Tissue kit (**Table 3.1**).

3.2.3 Primer Sequences For Marker Gene Amplification

All primer pairs utilised for the amplification of the five molecular markers Cytochrome *b* (Cyt *b*), 16S rRNA, NADH dehydrogenase subunit 4 (ND4), ribosomal S7 intron 1 (S7) and recombinase activator gene (RAG)2, investigated in this study, are summarised in **Table 3.2**. Sequences for both the external and internal primers for the partial amplification of the mitochondrial ND4 gene were designed using Primer Designer V1.01 (Scientific and Educational Software). Internal primers for the segmented amplification of the mitochondrial Cyt *b* gene were designed using LightCycler® Probe Design Software 2.0 (Roche Applied Sciences, Pty. Ltd.). All primer pairs were synthesized by the DNA Synthesis Laboratory, Department of Molecular and Cell Biology, University of Cape Town, South Africa. Stock solutions of 100 µM were prepared of each of the primers by calculating their extinction coefficients at 260 nm according to the following equation:

$$C_{primer} = A_{260} * \frac{100\ \mu M}{1.5(A)+0.71(C)+1.2(G)+0.84(T)}$$

Accordingly, working solutions of 20 µM were subsequently prepared from the stock solutions. All primers, stock solutions and working solutions were stored at -20°C until further use.

3.2.4 Amplification Of The Five Molecular Markers

Amplification of the five molecular markers was carried out individually (not multiplexed) using the Multiplex PCR Kit (Qiagen®) with their primer respective pairs as described in **Table 3.2**. In short, each PCR mixture contained the following: 2.2 µL RNase-free water, 1.5 µL (5X) Q-Solution, 1.3 µL (20 µM) forward primer, 1.3 µL (20 µM) reverse primer and 7.7 µL (2X) Qiagen® Multiplex PCR Master Mix. To this mixture, 1 µL of undiluted extracted DNA was added to obtain a final PCR reaction volume of 15 µL. Amplification was carried out according to the manufacturer's protocol consisting of a pre-denaturation step at 95°C for 15 min followed by 40 cycles of DNA amplification at 94°C for 30 seconds, 46°C - 56°C for 90 seconds and 72°C for 70-90 seconds, depending on the target sequence's length. A final extension completion period followed at 72°C for 10 minutes. Where necessary, samples were held at 15°C for no longer than one hour following completion of amplification. Samples were then stored at 4°C until further use. Amplification was performed in

either a Veriti™ 96-Well Thermal Cycler (Applied Biosystems, Pty. Ltd) or a 2720™ Thermal Cycler (Applied Biosystems, Pty. Ltd).

3.2.5 Electrophoresis And PCR Product Purification

All PCR products were analysed on a 1% agarose gel which was prepared as follows: 0.7 g of agarose powder was weighed off and added to 70 mL 1 X Casting buffer. This mixture was then gently swirled to obtain a homogenized solution which was then heated for two minutes at medium power in a microwave oven. The mixture was then allowed to cool to approximately 60°C (15 min) after which 7 µL (10,000X) GelRed™ Nucleic Acid Gel Stain (Biotium) was added. The liquid was then poured into a casting plate, the appropriate plastic comb/spacer inserted and a time period of 20-30 min allowed for the gel to solidify. The gel slab was then immersed in ± 600 mL of a 1 X Running buffer after which a sufficient volume of PCR product was then mixed with 1 µL loading buffer and loaded onto the 1% agarose gel. To estimate amplicon sizes, 5 µL of an appropriate molecular weight marker (GeneRuler™ DNA Ladder Mix) was then loaded onto the 1% agarose gel. Electrophoresis was then performed at 110 V for 65 min using an EPS 601 (Amersham Pharmacia Biotech, Inc.) electrophoresis setup. Visualisation of the PCR amplicons was done using an E-Box CN-1000 UV transilluminator and portable darkroom (Vilber Lourmat). The subsequent purification of the PCR amplicons was accomplished by means of the Wizard® SV Gel and PCR Clean-Up System (Promega) according to the manufacturer's instructions. Where required, the purified PCR amplicons were concentrated using a SpeedVac Concentrator (Sarvant Instruments, Farmingdale, N.Y.) after which 2 µL of each sample was analysed on a 1% agarose gel to determine DNA concentration. A detailed summary of the constituents of each of the buffers used during electrophoresis and PCR product purification are given in **Addendum A**.

3.2.6 Cycle Sequencing Of Purified PCR Fragments

The five molecular markers were bi-directionally sequenced using their respective primer pairs as summarised in **Table 3.2**. Each cycle sequencing reaction consisted of 2 µL Half-Dye Mix® (Bioline), 0.5 µL BigDye® (Terminator v3 Cycle Sequencing Kit, Applied Biosystems), 2 µL (0.8 µM) primer and 4.5 µL Milli-Q® water. One microliter of purified PCR product was then added to the sequencing mixture to obtain a final reaction volume of 10 µL. Cycle sequencing was performed using two seperate protocols. A standard three hour cycle sequencing protocol was used for the nuclear markers (S7 and RAG2) which involved 35 cycles of DNA amplification at 96°C for 10 seconds, 52°C for 30 seconds and 60°C for 4 minutes. A final extension completion period followed at 60°C for 10 minutes.

Table 3.2 A summary of the primers used in this study. Summarised are the primer names, primer sequences, primer lengths, literature sources, annealing temperatures, amplicon lengths and aligned lengths.

Molecular Marker	Primer name	Primer sequence (5'-3')	Length (bp)	Reference	Annealing temp (°C)	Amplicon length (bp)	Aligned Length (bp)
Cyt *b*	$F_{A/S}$: Cyt *b* Ext F	ACC AAT GAC TTG AAA AAC CAC CG	23	Bellstedt group (unpublished)	46	900	1119
	$R_{A/S}$: SteenVis Int R	AGG GAT GGA GCG AAG AAT GGC	21	This study			
	$F_{A/S}$: SteenVis Int F	CTA CCT TCA CAT CGG ACG AGG	21	This study	56	835	
	$R_{A/S}$: Cyt *b* Ext R	GCG TCC GGT TTA CAA GAC CG	20	Bellstedt group (unpublished)			
16S rRNA	$F_{A/S}$: 16Sa-L2510	CGC CTG TTT ATC AAA AAC AT	20	(Palumbi et al., 1991)	46	650	613
	R_A: 16Sb-H3080	CCG GTC TGA ACT CAG ATC ACG T	22				
	R_S: Chan16SF1	CAC TCT AAA ACA AAA GGC	18	This study			
ND4	$F_{A/S}$: ChanND4F3	CCC ATG GTA CAG ATC GAC	18	This study	52	1500	637
	R_A: ChanND4R6	CCT CTG TCT TTA GAA TCA C	19	This study			
	R_S: IntND4R1	CGT ATT ATA CCA TAG CC	17	This study			
S7	$F_{A/S}$: S7 RPEX1F	TGG CCT CTT CCT TGG CCG TC	20	(Chow and Hazama, 1998)	56	650	611
	$R_{A/S}$: S7 RPEX2R	AAC TCG TCT GGC TTT TCG CC	20				
RAG2	$F_{A/S}$: NeoRAG2F	AAA CTG AGG GCC ATT TCC TT	20	(López-Fernández et al., 2010)	55	950	838
	$R_{A/S}$:NeoRAG2R	GGG TTC TTT CTT CCT CTT TGG	21				

$F_{A/S}$, $R_{A/S}$: Forward/ Reverse primer used for amplification (A) and cycle sequencing (S).

R_A: Reverse primer used for amplification (A) only.

R_S: Reverse primer used for cycle sequencing (S) only.

Furthermore, the one hour ST_eP cycle sequencing protocol (Platt et al., 2007), was used for the three mitochondrial markers Cyt *b*, 16S rRNA and ND4. The protocol involved an initial incubation step at 96°C for 1 minute followed by 15 cycles of DNA amplification at 96°C for 10 seconds, 52°C for 5 seconds and 60°C for 75 seconds. Next, the elongation time was extended by 15 s for cycles 16-20 followed by an additional extension of 30 seconds for cycles 21-25. Cycle sequencing was carried out in a Veriti™ 96-Well Thermal Cycler (Applied Biosystems, Pty. Ltd) or a 2720™ Thermal Cycler (Applied Biosystems, Pty. Ltd). The sequencing products were analysed using an ABI® 3100 Genetic Analyser at the Central Analytical Facility, University of Stellenbosch, Western Cape, South Africa.

3.2.7 Nucleotide Sequence Analysis And Alignment

All subsequent chromatograms were edited using ChromasPro (v 1.5, Technelysium, Pty. Ltd.) and the nucleotide sequences exported as .txt files for analysis in BioEdit® Sequence Alignment Editor (v 7.1.3.0) (Hall, 1999). In addition, the nucleotide sequences of 154 Neotropical cichlids as incorporated by (López-Fernández et al., 2010), were downloaded from GenBank and are summarised in **Addendum B**. Five alignment matrices representing each of the five molecular markers were then generated. The six Neotropical cichlids *Chromidotilapia guntheri*, *Etroplus maculatus*, *Hemichromis fasciatus*, *Heterochromis multidens*, *Paratilapia polleni* and *Paretroplus polyactis* were used as the outgroups. Unfortunately, some of the outgroup species did not have representative nucleotide sequences for all of the five molecular marker genes (Cyt *b*, 16S rRNA, ND4, RAG2 and S7) available on GenBank at the time of this study. According to the supplementary data of López-Fernández et al. (2010), the GenBank accession number for the Cyt *b* nucleotide sequence of *Hemichromis fasciatus* was AY050618. However, when downloaded, the species was specified as *Nannacara anomala*. Hence, '*Hemichromis fasciatus'* was excluded from the Cyt *b* sequence matrix. All of the nucleotide sequences (reference and newly sequenced) were aligned in five separate matrices, each representing one of the five molecular markers (Cyt *b*, 16S rRNA, ND4, S7 and RAG2), by using the alignment tool Clustal W (v 1.4) (Thompson et al., 1994) available in the BioEdit® Sequence Alignment Editor (v 7.1.3.0) (Hall, 1999). Finer alignment of the five separate matrices was done by eye.

3.2.8 Phylogenetic Analysis

3.2.8.1 Character variability assessment and parsimony analysis

Parsimony analysis was performed using the software package PAUP* (v 4.0b10) (Swofford, 2003). Each of the five aligned sequence matrices were trimmed at the exact same base at both the 5' and 3' ends so as to exclude missing characters.The five aligned nucleotide matrices were exported from BioEdit® Sequence Alignment Editor (v 7.1.3.0) (Hall, 1999) as Nexus files for the subsequent parsimony analyses. First, each of the five aligned matrices were analysed separately for the purpose of assessing their character variability and comparing their individual gene trees. Indel characters were coded as ?-marks in the Nexus output files used for phylogenetic analyses. Parsimony analysis was

performed on three combined datasets; one contained the aligned and concatenated mitochondrial sequences (Cyt *b*, 16S rRNA and ND4), another contained the aligned and concatenated nuclear sequences (RAG2 and S7) and a third contained the aligned and concatenated mitochondrial and nuclear sequences. The latter thus formed the five-gene sequence matrix. In addition, the character variability of *Pterophyllum* nucleotide sequences was assessed for each of the five molecular markers by calculating the number of variable characters. Gaps were not calculated as a fifth character.

The heuristic search was set to complete 1000 replicates by using the Tree Bisection and Reconnection (TBR) branch swapping algorithm. Furthermore, instructions were given to keep and swap on best trees only and to save no more than 10 trees of score (length) greater and equal to five in each replicate. All trees were rooted to the Madagascan cichlid species *Paratilapia polleni* as used by López-Fernández et al. (2010). Upon the completion of the heuristic search the tree scores were computed. This entailed the total number of shortest trees retrieved per heuristic search in addition to the tree length, the consistency index (CI) and the retention index (RI). Thereafter, all the shortest trees were included in the computation of a strict consensus tree for each individual analysis. Finally, a bootstrap analysis consisting of a full heuristic search of 1000 replications was performed. Branches with values ≥75% were accepted to be well supported whereas values ranging between 74% and 50% were considered to be only moderately supported.

3.2.8.2 Maximum likelihood (ML) analysis

Each of the five aligned sequence matrices were trimmed at the exact same base at both the 5’ and 3’ ends so as to exclude missing characters. The five trimmed sequence matrices were then concatenated according to the same methodology applied for the parsimony analysis using the software package SequenceMatrix (Vaidya et al., 2011). Maximum likelihood analysis was performed using the tool RAxML-HPC2 on XSEDE version 8.0.24 (Stamatakis, 2014) available on The CIPRES Science Gateway V. 3.3 (Miller et al., 2010). The species *Paratilapia polleni* was chosen as the preferred outgroup. Model selection was performed using jModelTest 2.1.6 v20140903 (Darriba et al., 2012; Guidon and Gascuel, 2003). According to the AIC information criterion specified the GTR+I+G model for each partition. Each of the three focal data sets was partitioned using a mixed/partitioned model specifying each of the five sequence matrices as independent partitions. The analysis was allowed a maximum run time of six hours with the options to print branch lengths and to let RAxML halt bootstrapping automatically. At the end of the analysis, the output file shows the most likely tree with annotated bootstrap support values on nodes.

3.2.8.3 Generation of phylogenetic trees

The final tree recovered from each MP and ML analysis was displayed and analysed using the online tool *Interactive Tree Of Life* (iTOL) v 3.0 (Letunic and Bork, 2011, 2007). In setting the basic controls, the normal mode was selected and it was specified to not align leaf labels. The advanced

control settings involved specifying all bootstrap values >50 to be displayed as text above each branch. Under the display controls the font size factor was kept at its default value of 1.0 X and the tree scale factor adjusted to 0.8 X.

3.2.8.4 Statistical model-selection for beast analysis

Statistical selection of the best-fit model of nucleotide substitution for the BEAST analysis was performed using jModelTest 2.1.6 v20140903 (Darriba et al., 2012; Guidon and Gascuel, 2003). Parameters were kept at default values while a maximum run time of six hours were specified. All four information criterions (AIC, AICc, BIC and DT) specified the GTR+I+G model for the Cyt *b* dataset.

3.2.8.5 BEAST analysis

Bayesian MCMC analysis was performed using the software program BEAST v1.8.2 (Drummond and Rambaut, 2007; Suchard and Rambaut, 2009) run on XSEDE available on The CIPRES Science Gateway v3.3 (Miller et al., 2010). First, a BEAST XML input file containing the configured evolutionary model parameters for BEAST was generated using the graphical user-interface application BEAUti v1.8.2 (Drummond et al., 2012). A Nexus file containing the aligned Cyt *b* sequences of 180 Neotropical cichlid taxa was then imported into the BEAUti file and divided into four distinct taxon sets; Cichlinae, Etroplinae, Ptychochrominae and Pseudocrenilabrinae (**Addendum C**). Each of these taxon sets was specified as a monophyletic clade. Next, the evolutionary model was specified. The GTR Substitution Model was selected with base frequencies prompted to be estimated. The G+I Sites option was selected as the Site Heterogeneity Model. An uncorrelated lognormal relaxed clock was chosen as the preferred molecular clock model whilst the Speciation: Birth-Death Process (Gernhard, 2008) was selected as the tree prior under which a random starting tree was to be generated. The Birth-Death Process was selected as it takes species evolution and extinction into consideration.

An uniform prior distribution was assigned to the ‘time to most recent common ancestor’ (tmrca) for each of the four (Cichlinae, Etroplinae, Ptychochrominae and Pseudocrenilabrinae) Neotropical subfamilies based on the posterior molecular age estimates recovered by Friedman et al. (2013). The reason why this approach of secondary dating was used was because Friedman et al. (2013) had used 10 fossil calibration dates to date the radiation of the Cichlidae. Furthermore, this was in agreement with the dating of McMahan et al. (2013). A summary of these prior settings are presented in **Table 3.3**. The treeModel prior (ucld.mean) was assigned a lognormal distribution with an infinite upper bound, “initial value” of 6.4E-4 and lower bound of zero. In setting the MCMC options, the chain length was set at 50,000,000 generations with parameters to be logged every 1,000 generations. The subsequent log file was analysed using the program Tracer (MCMC Trace Analysis Tool) version 1.6.0 (Rambaut et al., 2014). The program TreeAnnotator version 1.8.2 was used to explore the tree file for the best supported tree. Once identified, the tree with the highest posterior probability was viewed in FigTree (Tree Figure Drawing Tool) version 1.4.2. Based on the stochastic nature of the

Bayesian MCMC analysis, it was repeated four times after which the log and tree files were combined using the program LogCombiner version 1.8.2. Both TreeAnnotator version 1.8.0 and LogCombiner version 1.8.2 are included in the BEAST version 1.8.2 package (Drummond et al., 2012).

TABLE 3.3 A summary of the uniform prior distribution settings for each of the four monophyletic taxon sets. These settings include the mean ages as well as the upper and lower bounds (95% HPD range) for each of the four Neotropical subfamilies as recovered by Friedman et al. (2013).

Taxon set	Mean age (Ma ago)	Upper bound (Ma ago)	Lower bound (Ma ago)
Cichlinae	29.2	34.8	25.5
Etroplinae	36.0	42.2	30.3
Ptychochrominae	38.2	46.4	31.7
Pseudocrenilabrinae	43.7	51.6	38.2

3.3 RESULTS

3.3.1 DNA EXTRACTION

DNA was successfully isolated from 20 of the 27 tissue samples excluding the 7 lyophilised samples received from Dr. Dirk Neumann. The DNA concentrations of the 20 samples are presented in **Table 3.1**. Considering that five molecular markers (16S rRNA, Cyt *b*, ND4, S7 and RAG2) had to be amplified and that the working solution for each of the lyophilised samples was limited to 20 µL, the DNA concentrations of these samples were not determined prior to analysis. Compared to the PCI-method, the DNeasy Blood and Tissue kit gave an average DNA yield of 28.3 µg/µL DNA per sample whereas the PCI-method gave an average DNA yield of 18.8 µg/µL DNA per sample. In addition, the DNeasy Blood and Tissue kit required less starting material than the PCI-method. The DNA concentration of sample 5 was not measurable. This measurement was taken during a time in which the Nano-Drop apparatus was due for maintenance and a re-calibration.

3.3.2 AMPLIFICATION AND GEL ELECTROPHORESIS OF THE FIVE MOLECULAR MARKERS

All primer pairs as summarised in **Table 3.2** were effective in the amplification of their respective molecular markers as their resulting amplicons were of the expected sizes as was specified in literature (also as shown in **Table 3.2**). Individual gene amplifications were all optimised for optimal annealing temperatures in order to eliminate multiple band formation. Negative controls were run during these optimisation processes and consistently lead to no amplification. Amplification product identity was confirmed by sequencing which is described hereafter as no positive controls were available from other sources. The gel electrophoresis results presented hereafter, serve as a visual indication of the amplification success reached for each of the five molecular markers. The gel electrophoresis results

for each of the five molecular markers (Cyt *b*, 16S rRNA, ND4, RAG2, S7) are presented in **Figures 3.2** to **3.7** respectively.

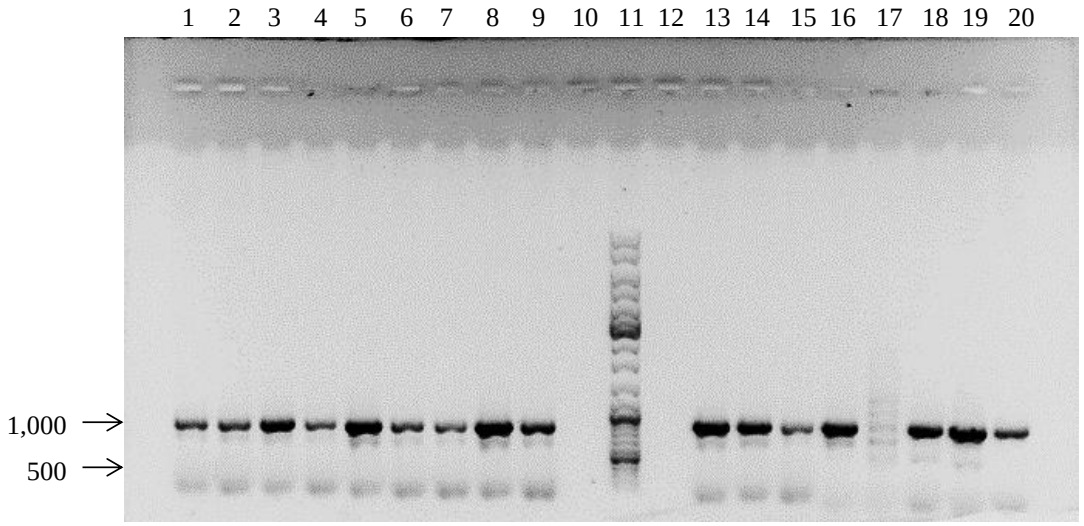


FIGURE 3.2: An image of a 1% agarose gel showing the separation of the "forward" and "reverse" PCR amplification products of the attempted amplification of the mitochondrial Cyt *b* gene. The Cyt *b* gene was amplified from 22 samples in total. The primer pair **Cyt *b* Ext F – SteenVis Int R** was used to generate the "forward" fragments of the Cyt *b* gene with expected amplification product size of 900 bp. In contrast, the primer pair **SteenVis Int F – Cyt *b* Ext R** was used to generate the "reverse" fragments of the Cyt *b* gene with expected amplification product size of 835 bp. The first 12 samples represent "forward" fragments of 900 bp. These are in lane 1: *Scalare* Guyana A, lane 2: *Scalare* Guyana B, lane 3: *Leopoldi* A, lane 4: *Scalare* Peru 1A, lane 5: *Scalare* Germany A, lane 6: *Scalare* Germany B, lane 7: *Altum* Rio Inirida 1, lane 8: *Altum* Rio Inirida 2, lane 9: *Altum* Rio Atabapo 1, lane 13: *Altum* Rio Atabapo 2, lane 14: *Altum* Rio Ventuari 3 and lane 15: *Altum* Rio Ventuari 4. The last five samples represent "reverse" fragments of 835 bp. These are in lane 16: *Scalare* Guyana A, lane 17: *Scalare* Guyana B, lane 18: *Scalare* Peru 1A, lane 19: *Scalare* Germany A and lane 20: *Altum* Rio Inirida 1. Lanes 10 and 12 were open whilst lane 11 contained the DNA ladder (GeneRuler™ DNA Ladder Mix).

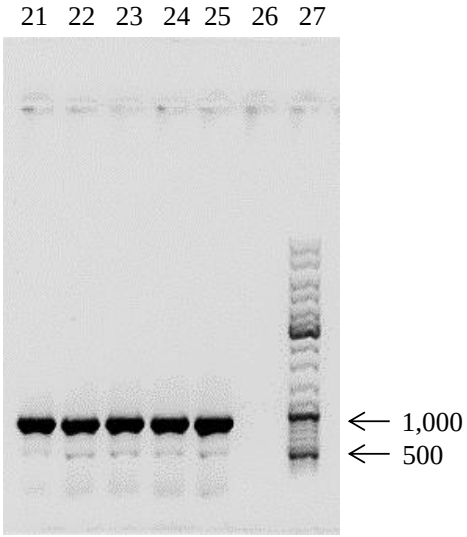


FIGURE 3.3: An image of a 1% agarose gel showing the separation of the remaining five "reverse" mitochondrial Cyt *b* PCR amplification products. The primer pair **SteenVis Int F – Cyt *b* Ext R** was used to generate the "reverse" fragments. The five samples amplified are in lane 21: *Altum* Rio Inirida 2, lane 22: *Altum* Rio Atabapo 1, lane 23: *Altum* Rio Atabapo 2, lane 24: *Altum* Rio Ventuari 3 and lane 25: *Altum* Rio Ventuari 4. Lane 26 was open and lane 27 contained the DNA ladder (GeneRuler™ DNA Ladder Mix).

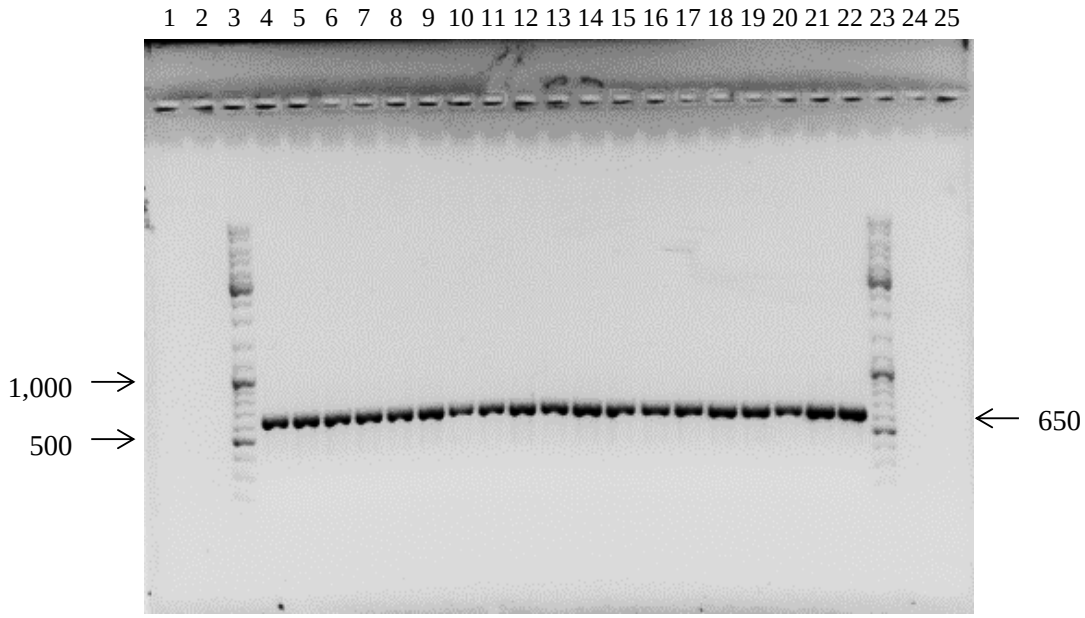


FIGURE 3.4: An image of a 1% agarose gel showing the separation of the PCR amplification products of the attempted amplification of the mitochondrial 16S rRNA gene. The primer pair **16Sa-L2510 – 16Sb-H3080** was used amplify the 16S rRNA gene from 16 samples. The expected amplification product size was 650 bp. Lanes 1 and 2 were open whilst lane 3 contained the DNA ladder (GeneRuler™ DNA Ladder Mix). The samples amplified were in lane 4: *Scalare* Guyana A, lane 5: *Scalare* Guyana B, lane 6: Santa Isabel A, lane 7: Santa Isabel B, lane 8: *Leopoldi* A, lane 9: *Leopoldi* B, lane 10: *Scalare* Peru 1A, lane 11: *Scalare* Peru 1B, lane 12: *Scalare* Germany A, lane 13: *Scalare* Germany B, lane 14: *Scalare* Xingu A, lane 15: *Altum* Rio Inirida 5, lane 16: *Altum* Rio Inirida 6, lane 17: *Altum* Rio Atabapo 1, lane 18: *Altum* Rio Atabapo 2, lane 19: *Altum* Rio Ventuari 3, lane 20: *Altum* Rio Ventuari 4, lane 21: *Altum* Rio Negro 1 and lane 22: *Altum* Rio Negro 2. Lane 23 contained the DNA ladder (GeneRuler™ DNA Ladder Mix) whilst lanes 24 and 25 were open.

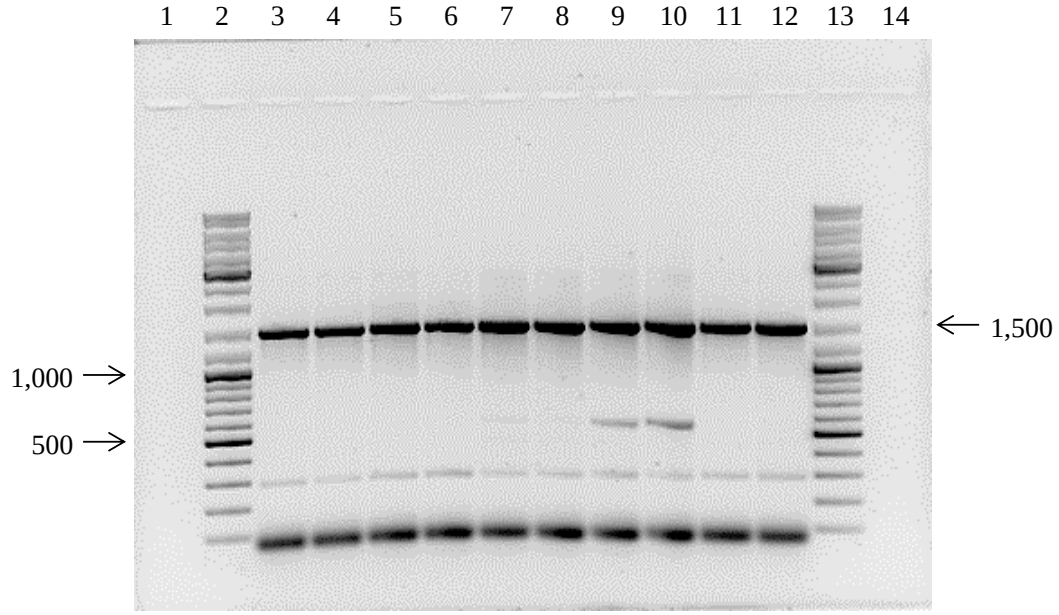


FIGURE 3.5: An image of a 1% agarose gel showing the separation of the PCR amplification products of the attempted amplification of the mitochondrial ND4 gene. The primer pair **ChanND4F3 – ChanND4R6** was used to amplify the ND4 gene from 10 samples. The expected amplification product size was 1,500 bp. Lane 1: was open, lane 2: DNA ladder (GeneRuler™ DNA Ladder Mix), lane 3: Scalare Germany D, lane 4: Scalare Germany E, lane 5: Scalare Xingu B, lane 6: Scalare Xingu C, lane 7: Scalare Manacapuru A, lane 8: Scalare Manacapuru B, lane 9: Leopoldi C, lane 10: Leopoldi D, lane 11: Scalare Germany F, lane 12: Scalare Germany G, lane 13: DNA ladder (GeneRuler™ DNA Ladder Mix) and lane 14: was open.

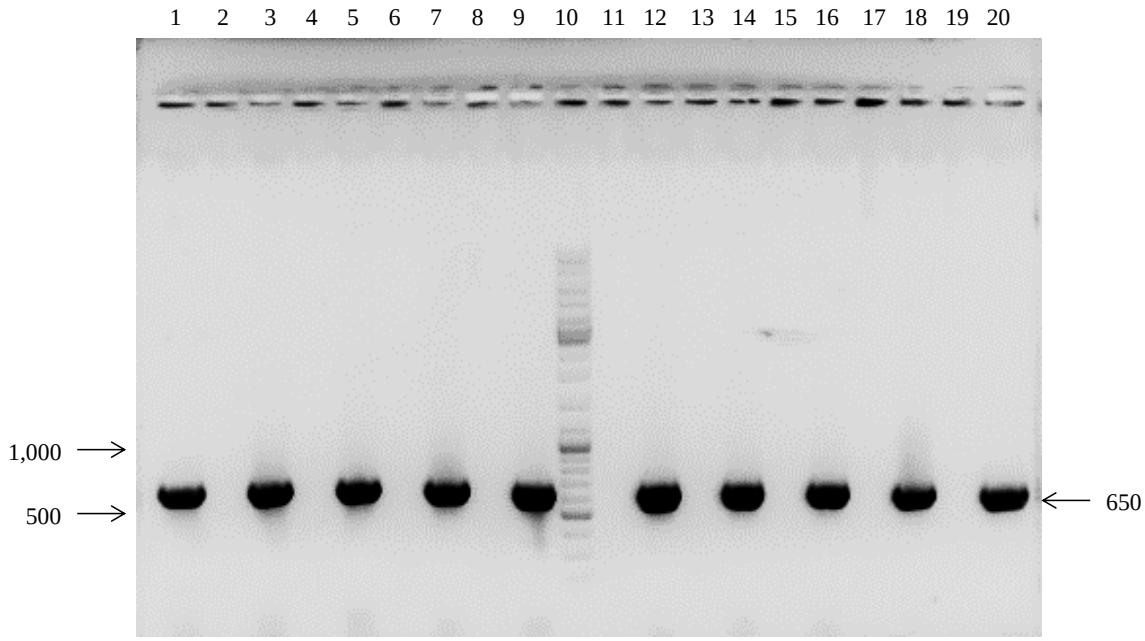


FIGURE 3.6: An image of a 1% agarose gel showing the separation of the PCR amplification products of the attempted amplification of the nuclear ribosomal S7 intron 1 region. The primer pair **S7 RPEX1F – S7 RPEX2R** was used to amplify the S7 intron 1 region from 10 samples. The expected amplification product size was 650 bp. Lane 1: *Scalare* Guyana A, lane 2: open, lane 3: *Scalare* Guyana B, lane 4: open, lane 5: Santa Isabel A, lane 6: open, lane 7: Santa Isabel B, lane 8: open, lane 9: *Leopoldi* A, lane 10: DNA ladder (GeneRuler™ DNA Ladder Mix), lane 11: open, lane 12: *Leopoldi* B, lane 13: open, lane 14: *Scalare* Peru 1A, lane 15: open, lane 16: *Scalare* Peru 1B, lane 17: open, lane 18: *Scalare* Germany A, lane 19: open and lane 20: *Scalare* Germany B.

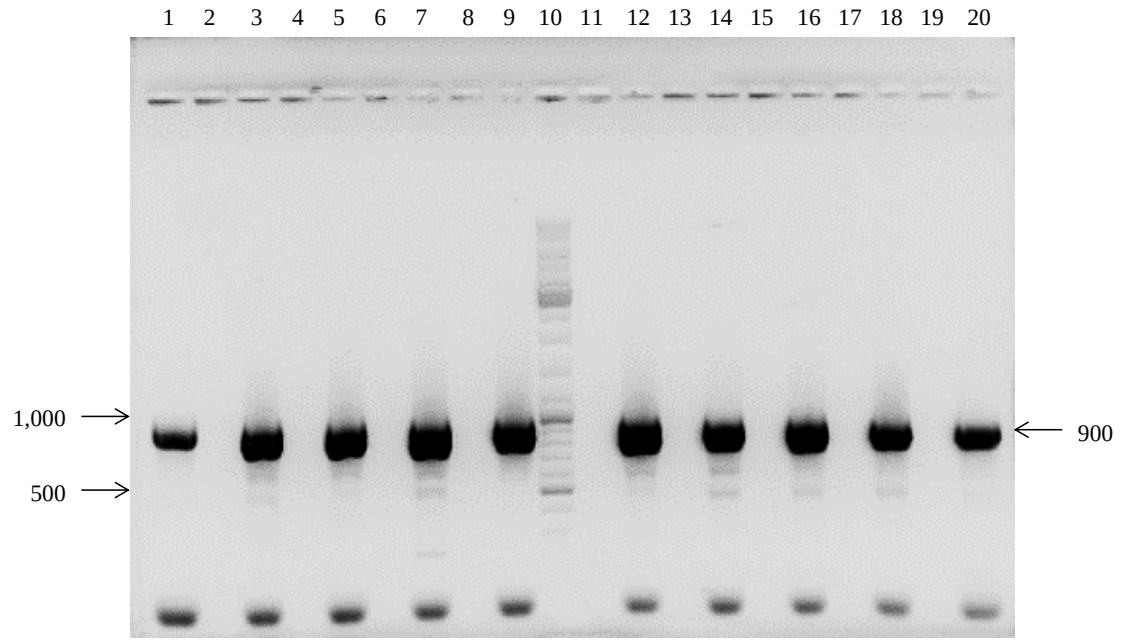


FIGURE 3.7: An image of a 1% agarose gel showing the separation of the PCR amplification products of the attempted amplification of the nuclear RAG2 gene. The primer pair **NeoRAG2F – NeoRAG2R** was used to amplify the RAG2 gene from 10 samples. The expected amplification product size was 900 bp. Lane 1: *Scalare* Guyana A, lane 2: open, lane 3: *Scalare* Guyana B, lane 4: open, lane 5: Santa Isabel A, lane 6: open, lane 7: Santa Isabel B, lane 8: open, lane 9: *Leopoldi* A, lane 10: DNA ladder (GeneRuler™ DNA Ladder Mix), lane 11: open, lane 12: *Leopoldi* B, lane 13: open, lane 14: *Scalare* Peru 1A, lane 15: open, lane 16: *Scalare* Peru 1B, lane 17: open, lane 18: *Scalare* Germany A, lane 19: open and lane 20: *Scalare* Germany B.

3.3.3 NUCLEOTIDE SEQUENCE ALIGNMENT AND ANALYSIS

3.3.3.1 Mitochondrial matrices

The aligned Cyt *b* sequence matrix of 1,119 bp consisted of 178 nucleotide sequences representative of five outgroup taxa, 153 Neotropical cichlid taxa and 20 *Pterophyllum* taxa. There was no need for the introduction of gaps in order to achieve optimal alignment. However, 11 of the Neotropical reference sequences (*Heterochromis multidens*, *Paretroplus polyactis*, *Paratilapia polleni*, *Acaronia nassa*, *Acaronia vultuosa*, *Apistogramma hoignei*, *Bioteocus dicentrarchus*, *Heroina isonycterina*, *Laetacara dorsigera*, *Parachromis friedrichsthalii*, *Santanoperca mapiritensis* and *Santanoperca pappaterra*) had significant amounts of missing data at their 3' ends. Furthermore, the aligned 16S rRNA sequence matrix of 613 bp consisted of 187 nucleotide sequences representative of six outgroup taxa, 165 Neotropical cichlid taxa and 22 *Pterophyllum* taxa. In total, 35 gaps had to be introduced into the 16S rRNA sequence matrix in order to achieve optimal alignment. Finally, the aligned ND4 sequence matrix of 637 bp consisted of 171 nucleotide sequences representative of four outgroup taxa, 150 Neotropical cichlid taxa and 21 *Pterophyllum* taxa. Six of the 150 reference sequences (*Pt.* scalare Genbank, *Cichla monoculus*, *Paratilapia polleni*, *Hemichromis fasciatus*, *Krobia sp*. and *Thorichthys affinis*) had significant amounts of missing data at either the 5' or 3' ends. There was no need for the introduction of gaps in order to achieve optimal alignment.

3.3.3.2 Nuclear matrices

The aligned RAG2 sequence matrix of 838 bp consisted of 173 nucleotide sequences representative of two outgroup taxa, 151 Neotropical cichlid taxa and 22 *Pterophyllum* taxa. Four of the 151 Neotropical reference sequences (*Andinoacara coeruleopunctatus*, *Crenicichla lenticulata*, *Petenia splendida* and *Theraps godmani*) had a significant amount of missing data (>10%) at their 3' ends. Furthermore, it was not necessary to introduce gaps into the sequence matrix in order to achieve its optimal alignment. In contrast, the aligned S7 sequence matrix of 611 bp consisted of 175 nucleotide sequences representative of five outgroup taxa, 154 Neotropical taxa and 21 *Pterophyllum* taxa. Two Neotropical reference sequences (*Paratilapia polleni* and *Paretroplus polyactis*) had a significant amount of missing data at the 3' ends. In addition, a significant number of gaps (more than 10% of aligned sequence length) had to be introduced in the matrix in order to achieve its optimal alignment.

3.3.3.3 Reduced five-gene sequence matrices

The aligned *Pterophyllum* sequences of the three mitochondrial (Cyt *b*, 16S rRNA and ND4) and two nuclear (RAG2 and S7) genes as well as the nucleotide sequences of seven Neotropical outgroup taxa (*Astronotus sp*., *Heros severus*, *Hoplarchus psittacus*, *Hypselecara coryphaenoides*, *Mesonauta insignis*, *Symphysodon aequifasciatus* and *Uaru fernandezyepezi*) are shown in **Addenda D**, **E**, **F, G** and **H** respectively. Although the 166 reference sequences of López-Fernández (**Addendum B**) were

included in the five sequence matrices, as described above, the reference sequences were excluded from the addenda as they have already been published and would have extended the thesis unnecessarily. Hence, the addenda serve to illustrate the nucleotide sequence variability of the three species of the genus *Pterophyllum* for each of the five phylogenetic markers.

3.3.4 PHYLOGENETIC ANALYSIS

3.3.4.1 Character variability assessment

A summary of the character variability, computed as a result of the heuristic search in a parsimony analysis, for each of the three aligned mitochondrial (Cyt *b*, 16S rRNA and ND4) sequence matrices is presented in **Table 3.4**.

TABLE 3.4 The character variability for each of the three aligned mitochondrial sequence matrices. Summarised are the total characters, the number of constant, parsimony uninformative (PU), parsimony informative (PI) and the total number of variable (PU+PI) characters for each of the three aligned mitochondrial sequence matrices.

	MITOCHONDRIAL MARKERS					
	16S rRNA		Cyt *b*		ND4	
Total characters	582	-	1119	-	637	-
Constant	284	48.8%	445	39.8%	148	23.2%
Parsimony Uninformative (PU)	57	9.8%	91	8.1%	56	8.8%
Parsimony Informative (PI)	241	41.4%	583	52.1%	433	68.0%
Variable (PU + PI)	298	51.2%	674	60.2%	489	76.8%

A summary of the character variability, computed as a result of the heuristic search in a parsimony analysis, for each of the two aligned nuclear (RAG2 and S7) sequence matrices is presented in **Table 3.5**.

TABLE 3.5 The character variability for each of the two aligned nuclear sequence matrices. Summarised are the total characters, the number of constant, parsimony uninformative (PU), parsimony informative (PI) and the total number of variable (PU+PI) characters for each of the three aligned mitochondrial sequence matrices.

	NUCLEAR MARKERS			
	RAG2		S7	
Total characters	838	-	611	-
Constant characters	452	53.9%	132	21.6%
Parsimony Uninformative (PU)	127	15.2%	98	16.0%
Parsimony Informative (PI)	259	30.9%	381	62.4%
Variable characters (PU + PI)	386	46.1%	479	78.4%

A summary of the character variability for the concatenated mitochondrial and nuclear sequence matrices is presented in **Table 3.6**.

TABLE 3.6 The character variability for the concatenated and aligned five-gene sequence matrix. Summarised are the total characters, the number of constant, parsimony uninformative (PU), parsimony informative (PI) and the total number of variable (PU+PI) characters for each of the three aligned mitochondrial sequence matrices.

	Character variability assessment				
	Mitochondrial (Cyt *b*, 16S rRNA, ND4)	Nuclear (RAG2, S7)	Total	% contribution	
				M	N
Total characters	2338	1449	3787	61.7	38.3
Constant characters	877	584	1461	60.0	40.0
Parsimony Uninformative (PU)	204	225	429	47.6	52.4
Parsimony Informative (PI)	1257	640	1897	66.3	33.7
Variable characters (PU + PI)	1461	865	2326	62.8	37.2

The three mitochondrial sequence matrices and the two nuclear sequence matrices therefore contributed 62.8% and 37.2% of the variable characters of the concatenated five-gene (Cyt *b*, 16S rRNA, ND4, RAG2 and S7) matrix respectively. Furthermore, of the 1,461 constant base pairs, the concatenated mitochondrial sequence matrix contributed 60.0% in contrast to the 40.0% of the concatenated nuclear sequence matrix. In addition, the concatenated mitochondrial sequence matrix contained 1,461 variable characters which represented 62.8% of the total variable characters observed

in the five-gene sequence matrix. In contrast, the concatenated nuclear sequence matrix could only account for 37.2% of the total variable characters observed in the five-gene sequence matrix.

A summary of the tree scores for each of the five molecular markers, the concatenated mitochondrial, concatenated nuclear and combined mitochondrial and nuclear matrices are presented in **Table 3.7**.

TABLE 3.7 A summary of the tree scores for each of the eight sequence matrices analysed by maximum parismony. Summarised is the total number of trees, the length of the first tree of all the shortest possible trees and the CI and RI values for each of the eight sequence matrices analysed.

Tree scores	Cyt *b*	16S rRNA	ND4	RAG2	S7	Mito	Nuc	M+N
Total trees	131	200	60	6,650	8,400	30	6,480	1,430
Length of 1# tree	10,179	2,249	6,918	908	1,630	19,877	2,607	22,655
CI	0.122	0.231	0.143	0.561	0.534	0.138	0.529	0.182
RI	0.599	0.691	0.593	0.844	0.822	0.596	0.820	0.624

The results of the character variability assessment for the three mitochondrial and two nuclear aligned sequence matrices, containing only the species of the genus *Pterophyllum*, are presented in **Tables 3.8** and **3.9** respectively. Thereafter a summary of the character variability of the combined mitochondrial and nuclear sequence matrix is presented in **Table 3.10** which reflects the character variability of the aligned and concatenated five-gene (Cyt *b*, 16S rRNA, ND4, RAG2 and S7) sequence matrix. Finally, a summary of the tree scores for each of the five *Pterophyllum* matrices is presented in **Table 3.11**.

TABLE 3.8 The character variability in the three aligned mitochondrial sequence matrices of the genus *Pterophyllum*. Summarised are the total characters, the number of constant, parsimony uninformative (PU), parsimony informative (PI) and the total number of variable (PU+PI) characters for each of the three aligned mitochondrial sequence matrices.

	MITOCHONDRIAL MARKERS					
	16S rRNA		Cyt *b*		ND4	
Total characters	582	-	1119	-	637	-
Constant characters	547	94.0%	935	83.6%	526	82.6%
Parsimony Uninformative (PU)	1	0.2%	4	0.4%	9	1.4%
Parsimony Informative (PI)	34	5.8%	180	16.1%	102	16.0%
Variable characters (PU + PI)	35	6.0%	184	16.4%	111	17.4%

TABLE 3.9 The character variability in the two aligned nuclear sequence matrices of the genus *Pterophyllum*. Summarised are the total characters, the number of constant, parsimony uninformative (PU), parsimony informative (PI) and the total number of variable (PU+PI) characters for each of the three aligned mitochondrial sequence matrices.

	NUCLEAR MARKERS			
	RAG2		S7	
Total characters	838	-	611	-
Constant characters	821	98.0%	592	96.9%
Parsimony Uninformative (PU)	0	0	2	0.3%
Parsimony Informative (PI)	17	2.0%	17	2.8%
Variable characters (PU + PI)	17	2.0%	19	3.1%

TABLE 3.10 The combined mitochondrial and nuclear gene character variability of the genus *Pterophyllum*. Summarised are the total characters, the number of constant, parsimony uninformative (PU), parsimony informative (PI) and the total number of variable (PU+PI) characters for each of the three aligned mitochondrial sequence matrices.

	Character variability assessment				
	Mitochondrial (Cyt *b*, 16S rRNA, ND4)	Nuclear (RAG2, S7)	Total	% contribution	
				M	N
Total characters	2338	1449	3787	61.7	38.3
Constant characters	2008	1413	3421	58.7	41.3
Parsimony Uninformative (PU)	14	2	16	87.5	12.5
Parsimony Informative (PI)	316	34	350	90.3	9.7
Variable characters (PU + PI)	330	36	366	90.2	9.8

TABLE 3.11 A summary of the tree scores for each of the five *Pterophyllum* sequence matrices.

Sequence matrix	Number of trees	Length of first tree	CI	RI
Cyt *b*	1	190	0.979	0.995
16S rRNA	1	37	0.973	0.994
ND4	11	123	0.943	0.986
RAG2	1	19	0.895	0.935
S7	1	19	1.000	1.000

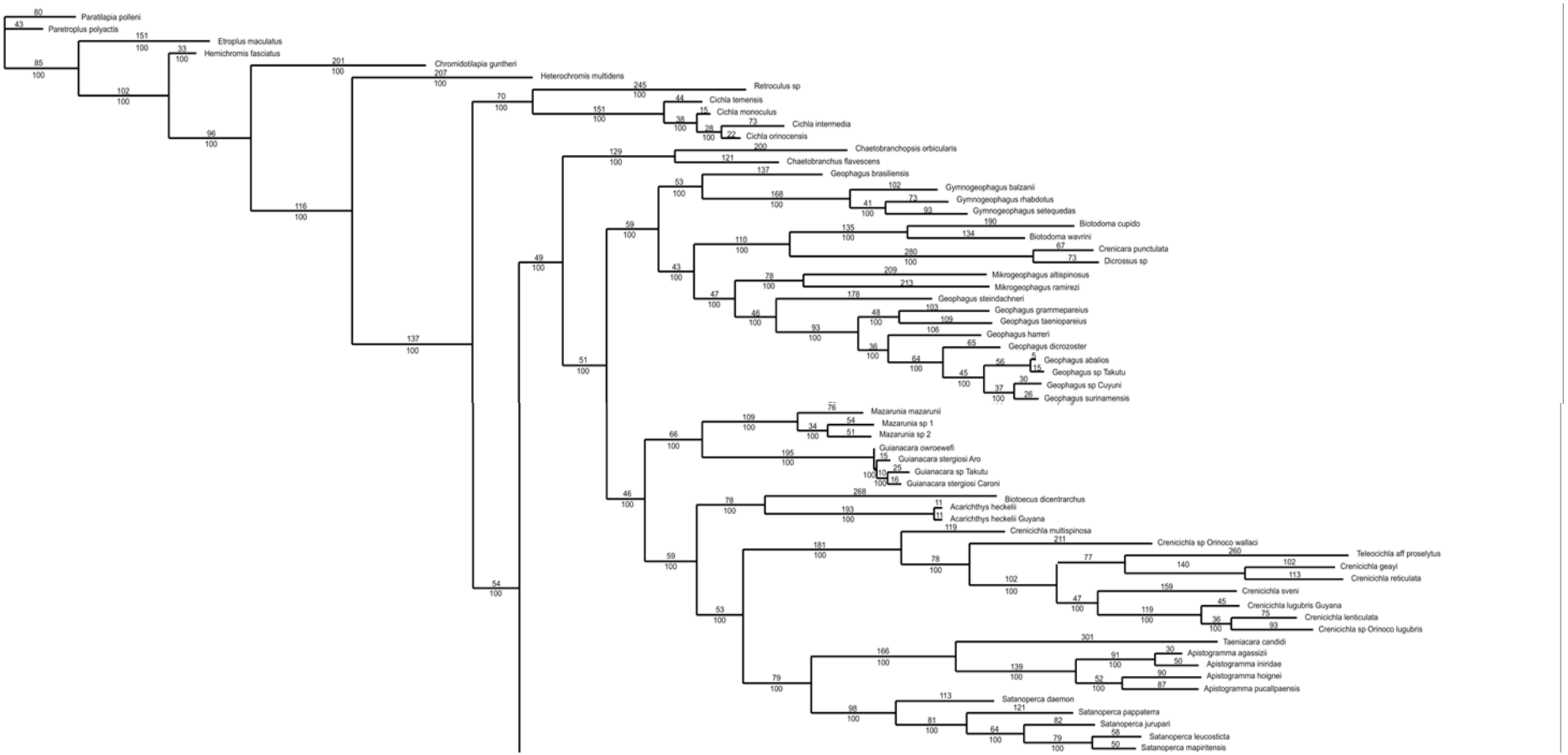
3.3.4.2 Parsimony (MP) analysis

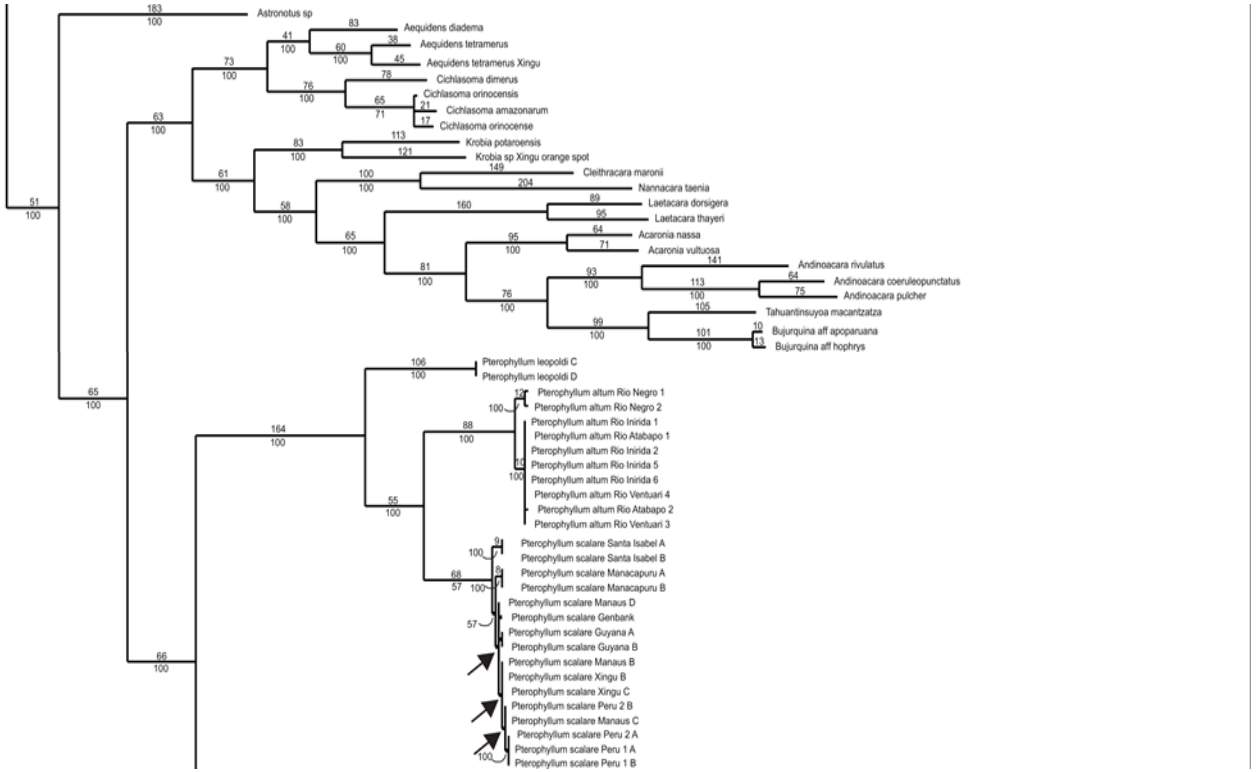
Parsimony analyses were performed on the five individual sequence matrices as well as the concatenated mitochondrial, concatenated nuclear and five-gene sequence matrix. However, the tree topologies retrieved for these trees were very similar and in some instances identical to the tree topologies retrieved from the ML analyses (presented next) performed on the same sequence matrices. For this reason, only the tree topology retrieved for the five-gene sequence matrix is presented in **Figure 3.8**.

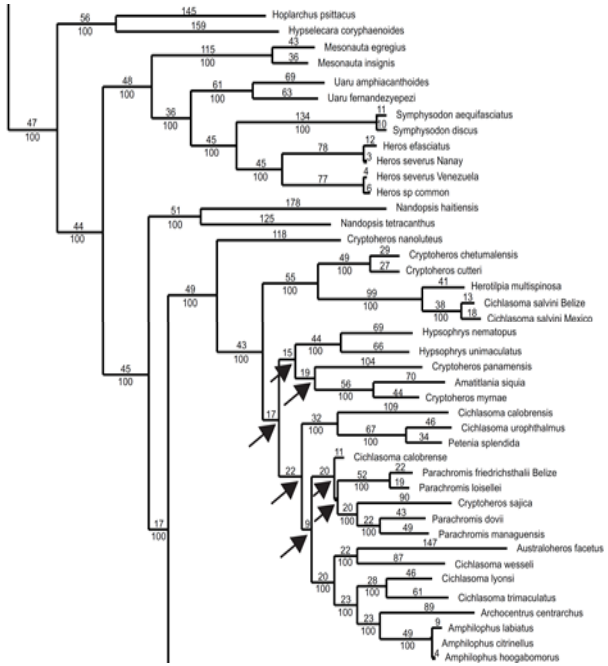
3.3.4.3 Maximum likelihood (ML) analysis

The three (Cyt *b*, 16S rRNA and ND4) mitochondrial gene trees are presented as **Figures 3.9** to **3.11** respectively. In spite of varying bootstrap support, the topology of the genus *Pterophyllum* remained the same throughout the three gene trees with the only difference among the trees being the placement of the genus with regard to its immediate neighbouring Heroini genera, more specifically the South American species *Hoplarchus psittacus*, *Hypselecara coryphaenoides* and the Mesonautines. **Figure 3.12** shows the combined mitochondrial (Cyt *b*, 16S rRNA and ND4) gene tree in which the genus *Pterophyllum* was retrieved as a monophyletic clade consisting of the three species *Pt. leopoldi*, *Pt. altum* and *Pt. scalare.* Of these three species *Pt. leopoldi* was recovered as the basal lineage and sister to the two more closely related *Pt. altum* and *Pt. scalare* clades. Furthermore, the genus was recovered as the basal lineage of the tribe Heroini as described by López-Fernández et al. (2010).

The two (RAG2 and S7) nuclear gene trees are presented in **Figures 3.13** and **3.14** respectively. Both the RAG2 and S7 gene trees recovered the genus *Pterophyllum* as a monophyletic clade consisting of its three species *Pt. leopoldi*, *Pt. altum* and *Pt. scalare*. However, in context of the RAG2 gene tree, the two *Pt. leopoldi* representatives were recovered as the only fully supported clade. The *Pt. altum* and *Pt. scalare* groups were retrieved as a polytomy. In contrast, the S7 gene tree recovered all three species as well supported and separate clades. However, the *Pt. leopoldi* and *Pt. altum* clades were retrieved in a sister association and *Pt. scalare* as the basal lineage. The combined nuclear gene tree is presented as **Figure 3.15**. In this tree the genus *Pterophyllum* was recovered as a monophyletic clade consisting of the three species *Pt. leopoldi*, *Pt. altum* and *Pt. scalare*. The *Pt. leopoldi* representatives were recovered as a fully supported basal clade whilst the well supported *Pt. altum* and *Pt. scalare* clades were retrieved in a sister association. The representative taxa of both the *Pt. altum* and *Pt. scalare* clades formed polytomies. Furthermore, the genus *Pterophyllum* was recovered as the sister clade to the clade containing the Mesonautines and Central American Heroines. The two South American species *Hoplarchus psittacus* and *Hypselecara coryphaenoides* were retrieved in a sister association which formed the basal lineage of the tribe Heroini.







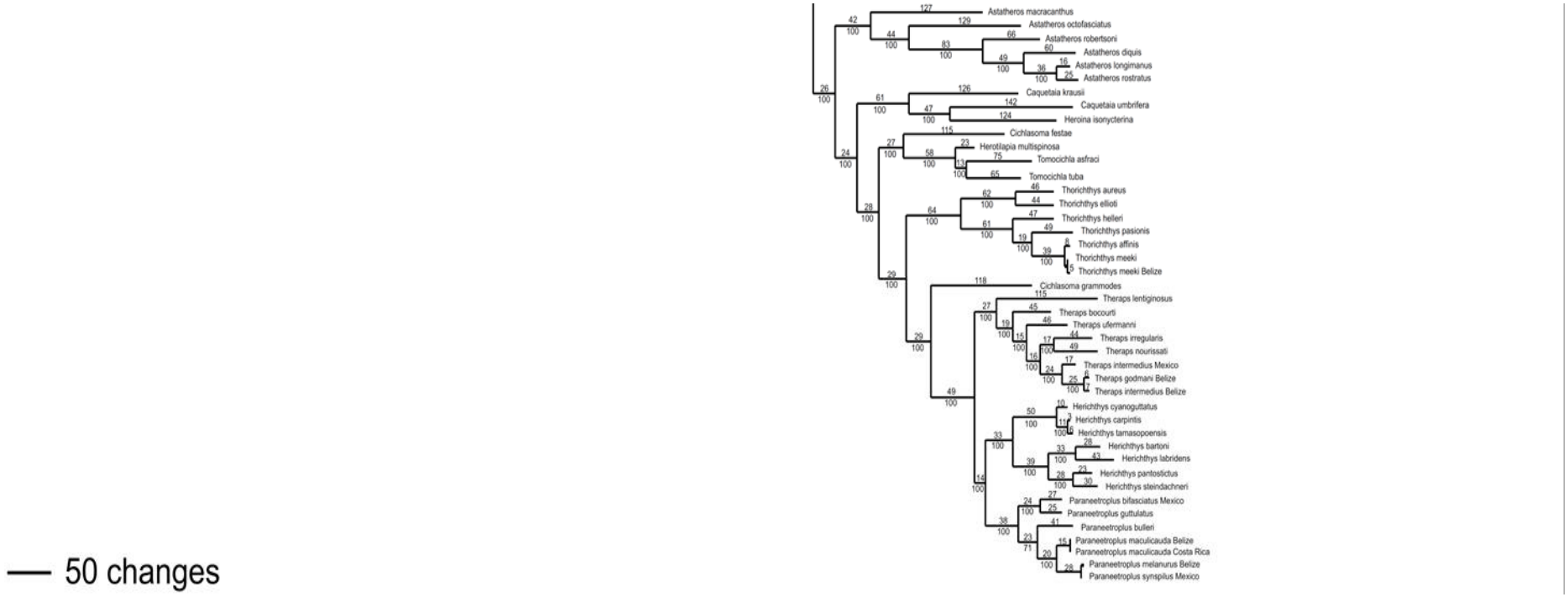
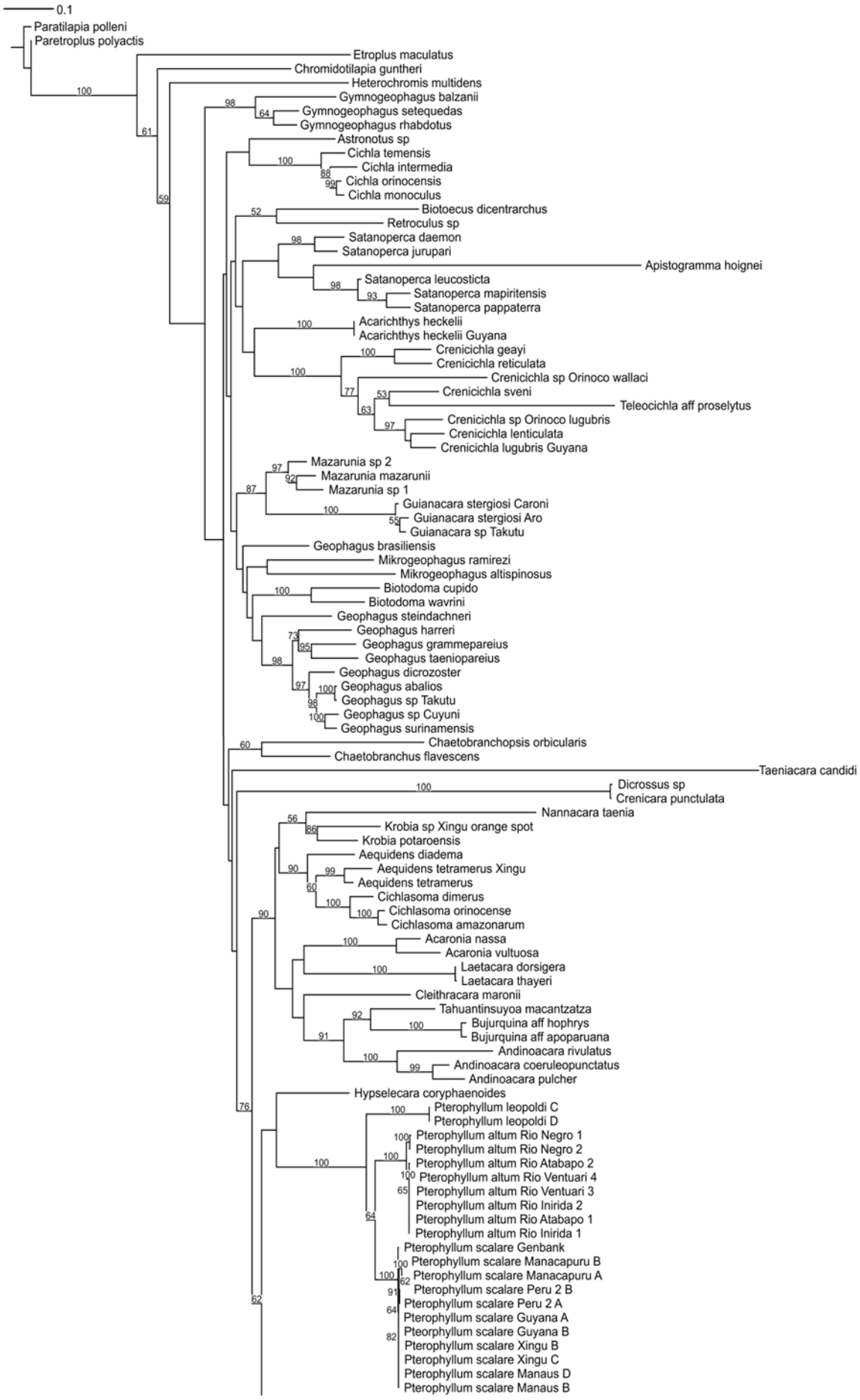


FIGURE 3.8: A changes tree generated by parsimony analysis performed on the concatenated and aligned five-gene (Cyt *b*, 16S rRNA, ND4, RAG2 and S7) sequence matrix. Branches are scaled proportionally and branch lengths are indicated above branches. Only bootstrap support values equal and greater than 50% are indicated below branches. Arrows indicate nodes which collapse in the strict consensus tree.

The concatenated five-gene (Cyt *b*, 16S rRNA, ND4, RAG2 and S7) species tree is presented in **Figure 3.16**. In this tree the genus *Pterophyllum* was recovered as a monophyletic clade consisting of three fully supported clades of which the *Pt. leopoldi* clade was recovered in a basal position. The clade consisting of the sister *Pt. altum* and *Pt. scalare* clades was well supported. The *Pt. altum* clade was subdivided into two well supported sister clades, the Rio Negro and Rio Orinoco clades, with bootstrap support values of 99% and 89% respectively. The Rio Orinoco clade was recovered as a polytomy. Likewise, the *Pt. scalare* representatives from Guyana, Rio Xingu and Manaus were recovered as a polytomy. In contrast, the *Pt. scalare* representatives from Peru, Santa Isabel and Lago Manacapuru were each recovered as a monophyletic clade. The Peruvian clade was weakly supported (<50%) whilst the Santa Isabel and Lago Manacapuru clades were well supported and were assigned bootstrap support values of 95% and 100% respectively. Furthermore, the two South American species *Hoplarchus psittacus* and *Hypselecara coryphaenoides* were recovered as the basal lineage of the tribe Heroini. The *Pterophyllum* clade was recovered as the sister clade of the clade containing the closely related monophyletic Mesonautines and Central American Heroine clade.



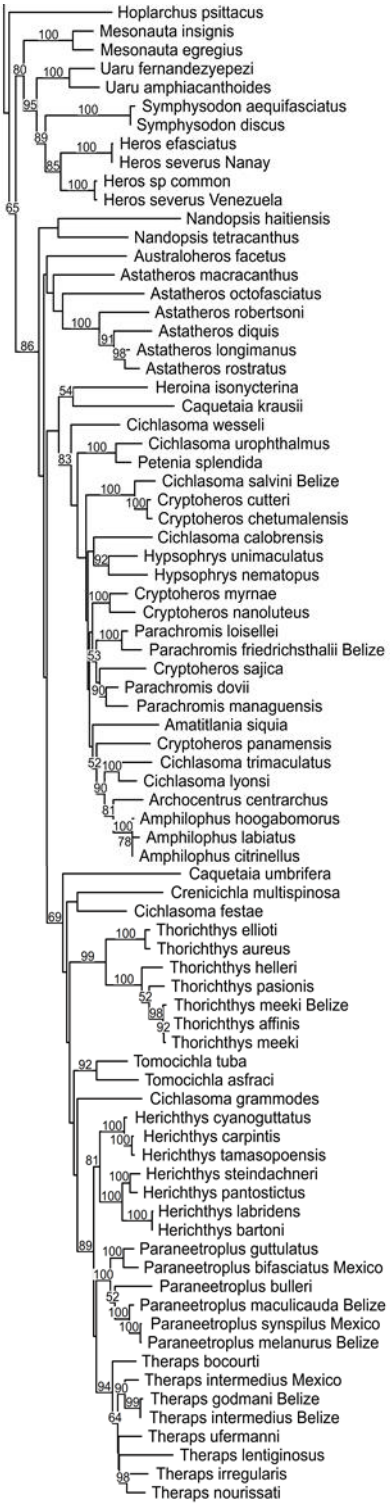


FIGURE 3.9: A phylogeny based on the aligned Cyt *b* sequence matrix as retrieved from maximum likelihood analysis. Branches are scaled proportionally and only bootstrap support values equal and greater than 50% are annotated above branches.

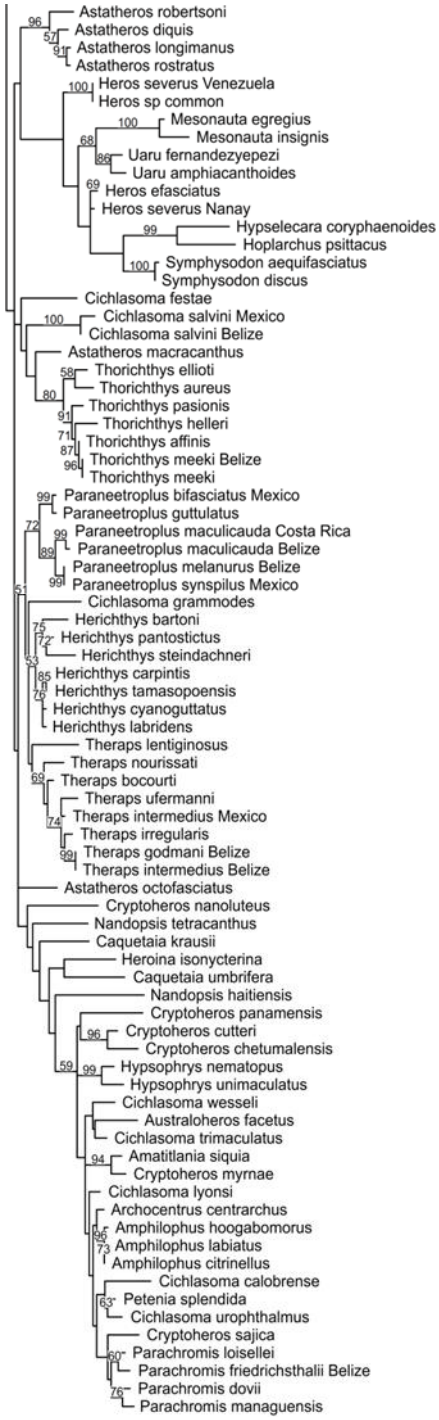


FIGURE 3.10: Maximum likelihood analysis performed on the aligned 16S rRNA sequence matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.

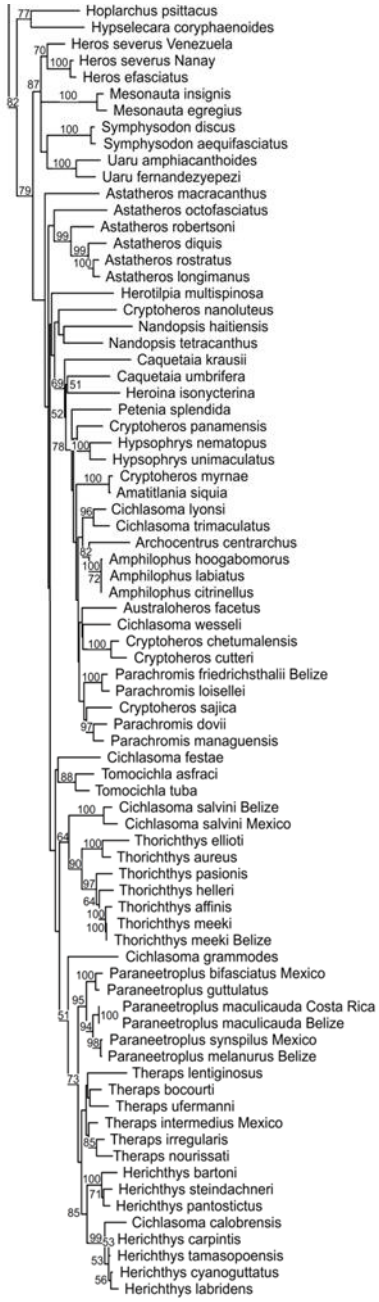
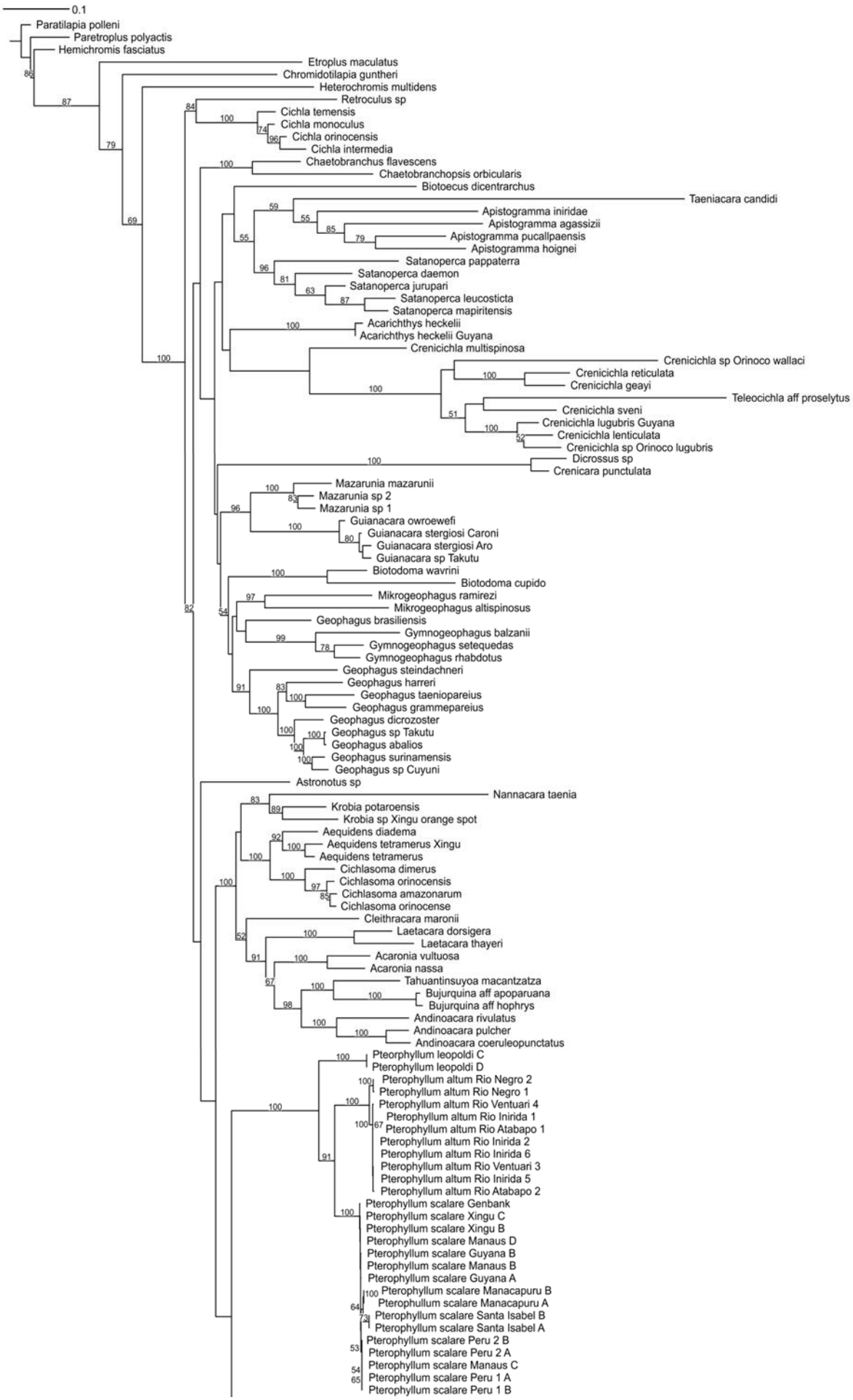


FIGURE 3.11: Maximum likelihood analysis performed on the aligned ND4 sequence matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.



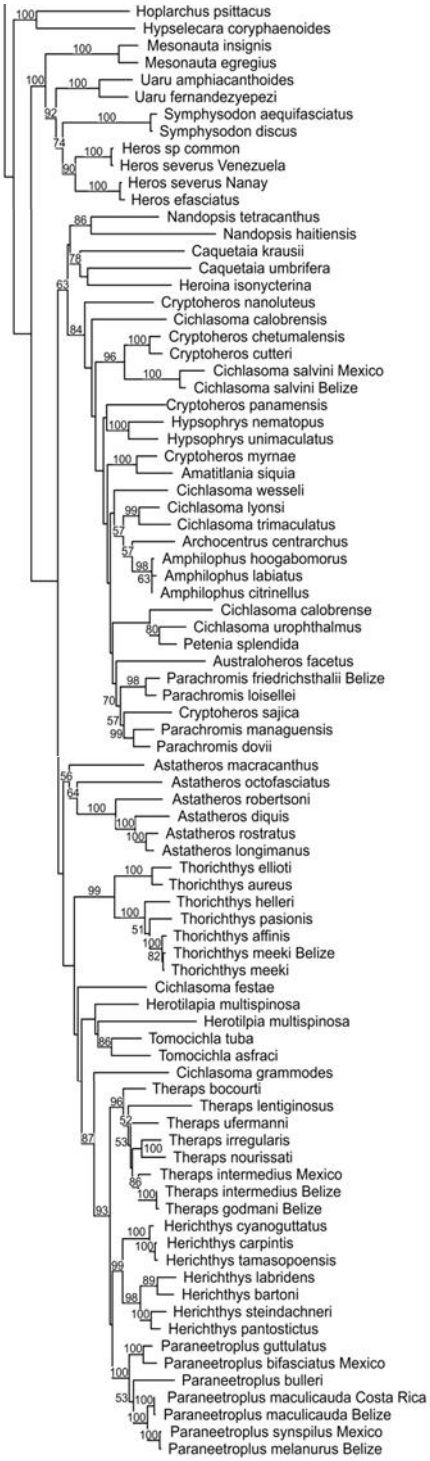
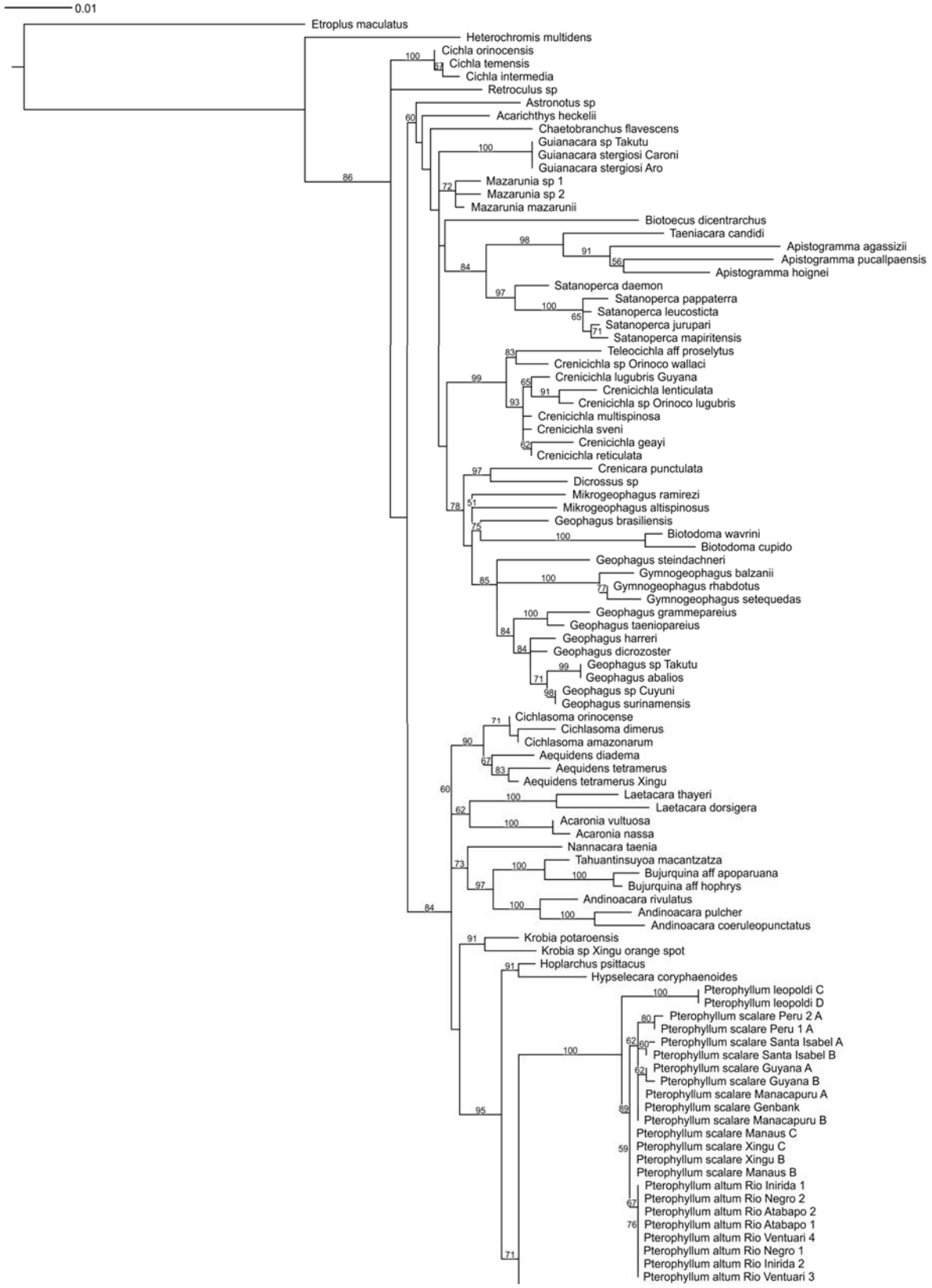


FIGURE 3.12: Maximum likelihood analysis performed on the aligned and concatenated mitochondrial (Cyt b, 16S rRNA, ND4) sequence matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.



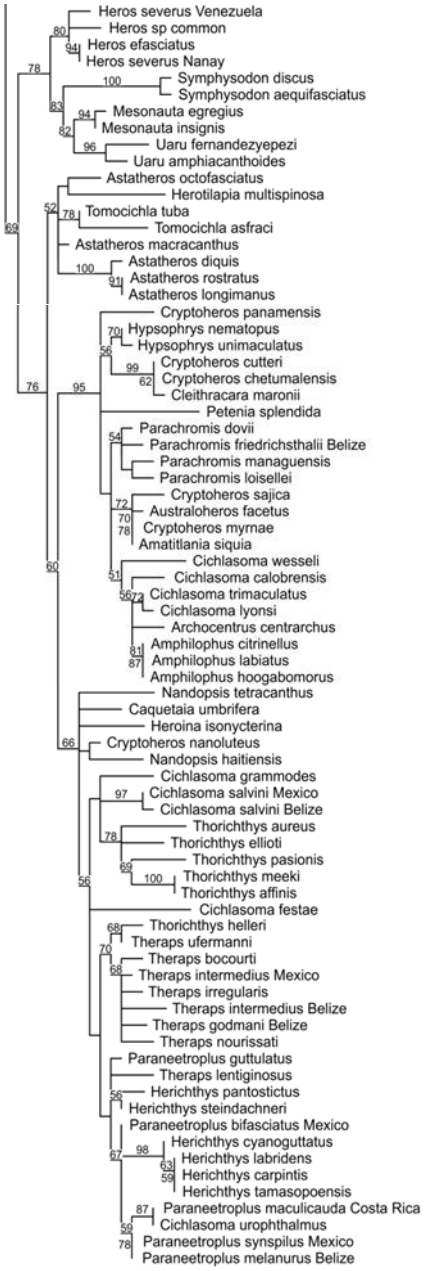
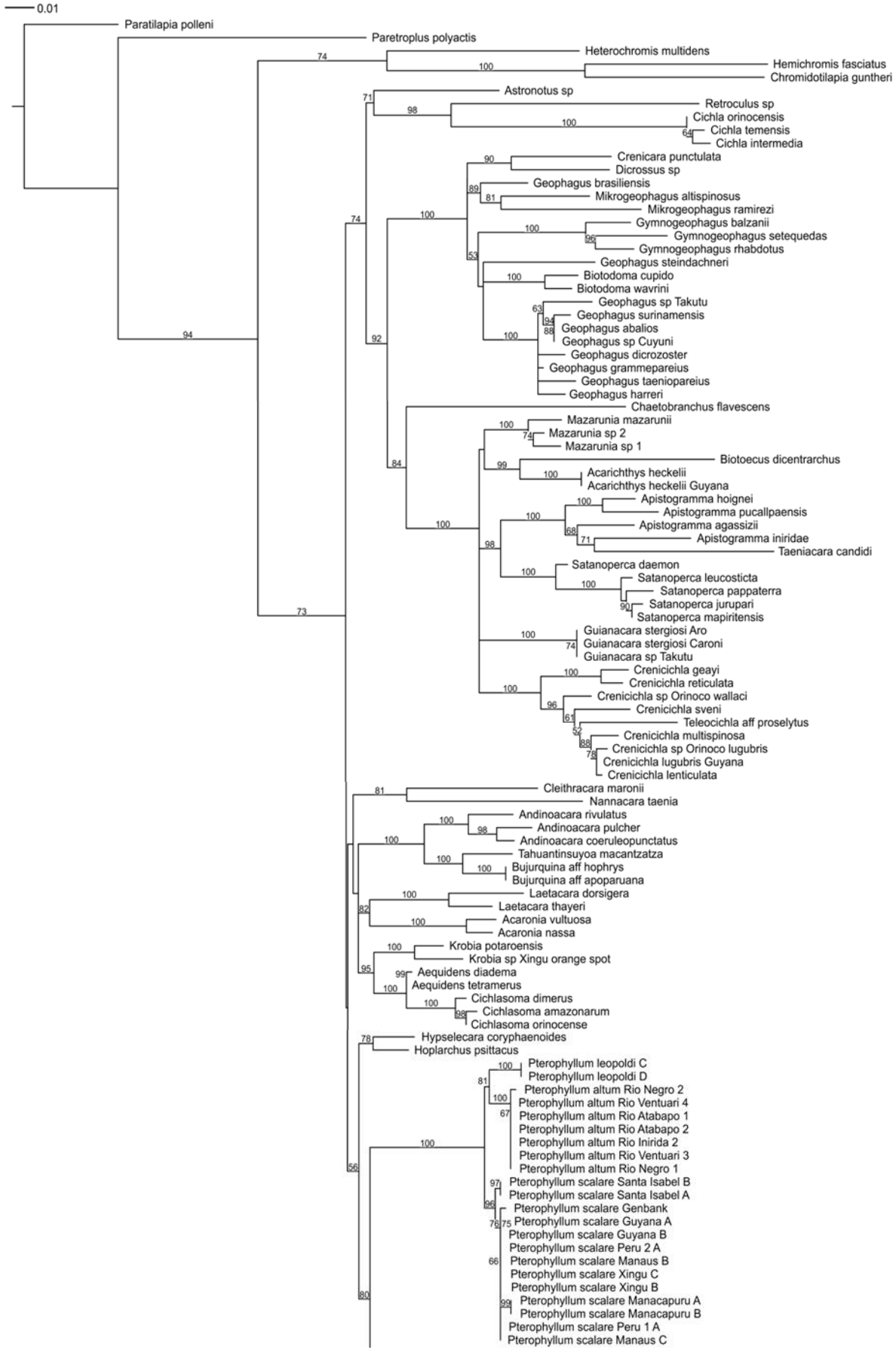


FIGURE 3.13: Maximum likelihood analysis performed on the aligned RAG2 sequence matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.



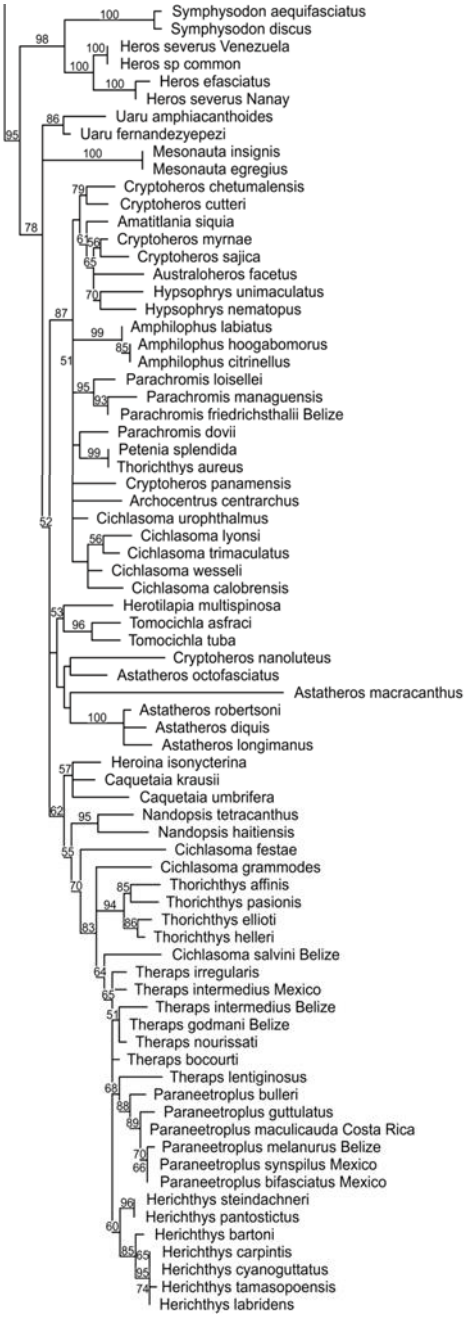
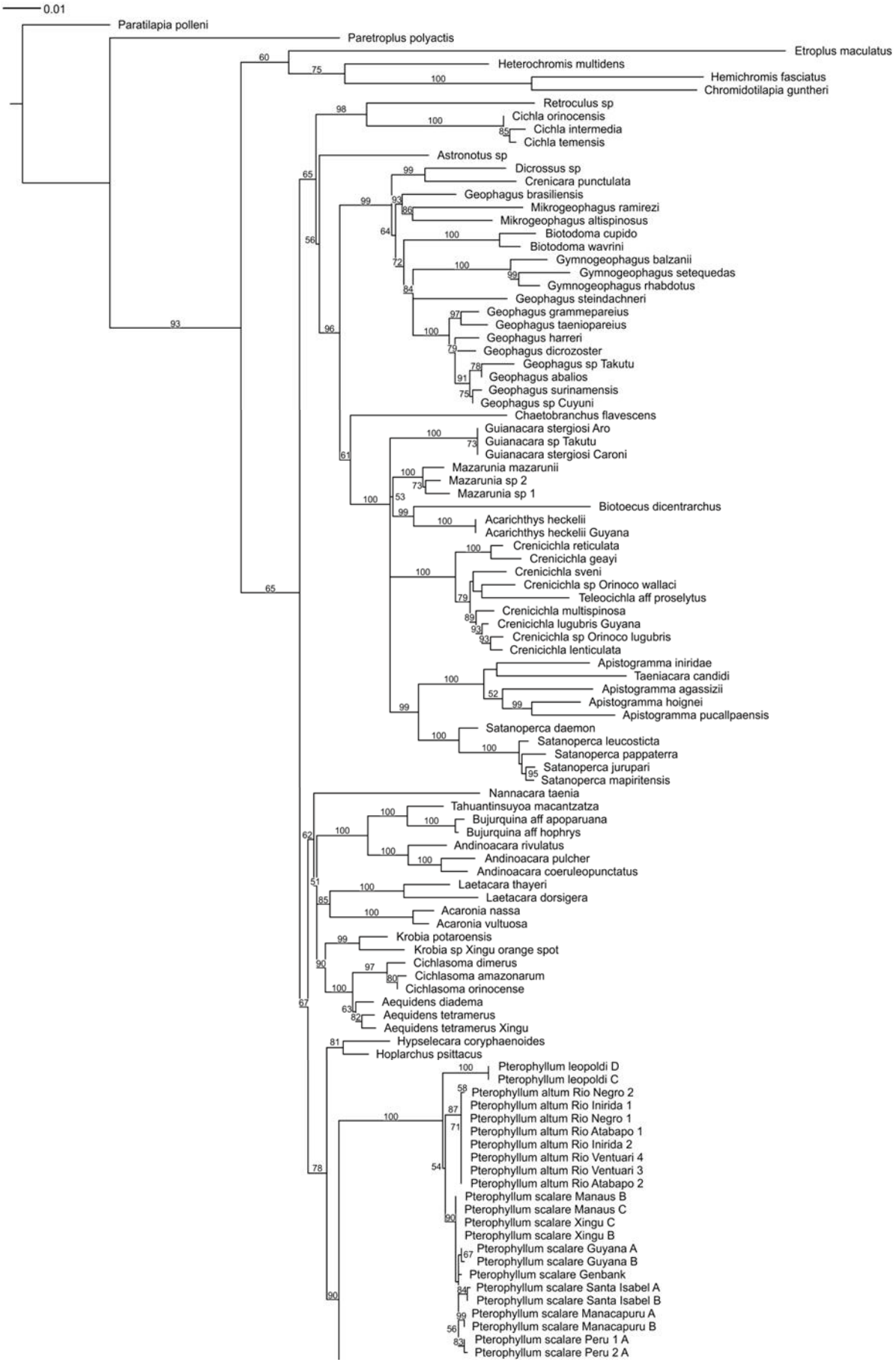


FIGURE 3.14: Maximum likelihood analysis performed on the S7 sequence matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.



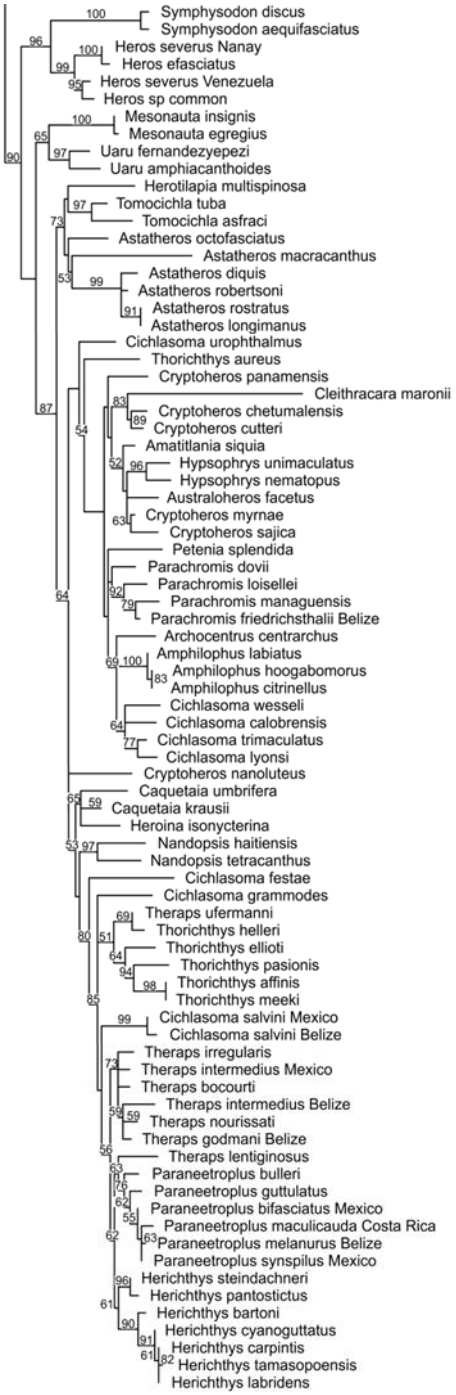
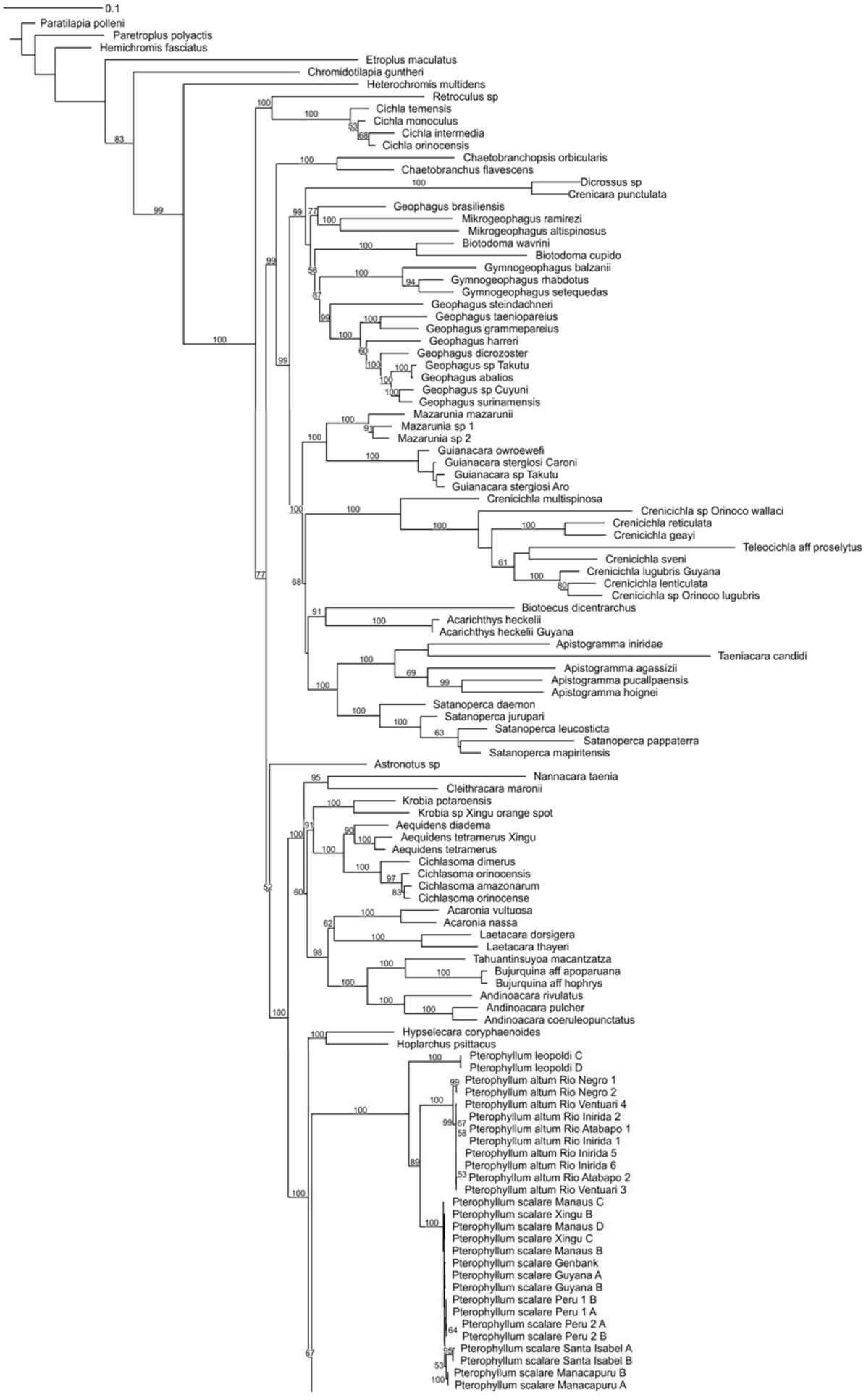


FIGURE 3.15: Maximum likelihood analysis performed on the aligned and concatenated nuclear (RAG2, S7) sequence matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.



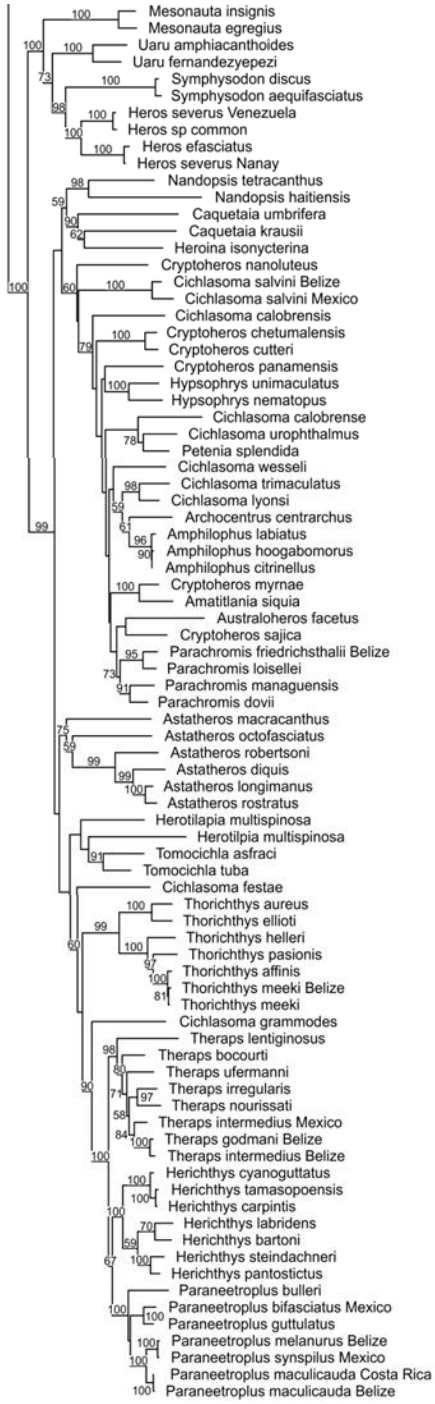


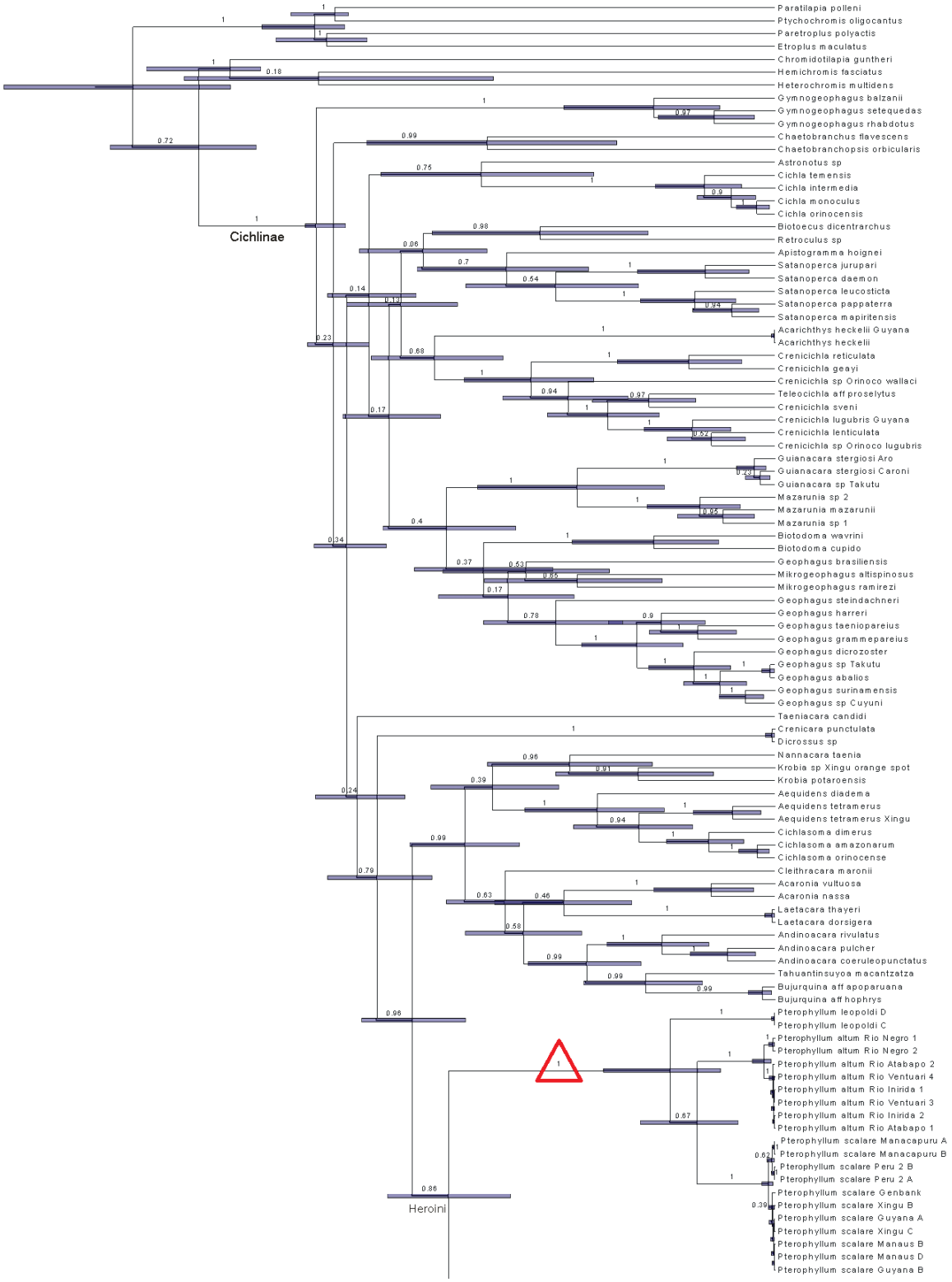
FIGURE 3.16: Maximum likelihood analysis performed on the aligned and concatenated five-gene (Cyt *b*, 16S, ND4, RAG2, S7) sequence matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.

3.3.4.4 BEAST analysis

The Bayesian analysis as performed on the Cyt *b* sequence matrix (**Figure 3.17**) recovered a basal lineage in the tribe Heroini consisting of the genus *Pterophyllum* nested with the two South American genera *Hoplarchus* and *Hypselecara*. The Mesonautines were recovered as a monophyletic group and sister to the Central American Heroines. This result is in accordance with the result obtained from the maximum likelihood analysis performed on the same data set. A summary of the posterior probabilities, branch times (Ma), mean node ages (Ma) and 95% HPD intervals for each of the clades are presented in **Table 3.12**.

TABLE 3.12 A summary of the temporal divergence of the genus *Pterophyllum* as recovered from the Bayesian analysis performed on the Cyt *b* sequence matrix. Results are summarised in terms of the posterior probability, branch time, mean node age and 95% highest posterior density (HPD) interval recovered for each clade.

Clades	Posterior probability	Branch times (Ma)	Mean node age (Ma)	95% HPD interval (Ma)
Pterophyllum	1	16.47	24.24	19.7 – 28.7
Pt. leopoldi	1	7.74	7.77	4.06 – 12.66
MRCA (*Pt. altum* & *Pt. sclare*)	0.67	2.00	7.77	4.06 – 12.66
Pt. altum	1	4.98	5.77	2.74 – 9.94
Rio Orinoco	1	0.63	0.79	0.3 – 1.61
Rio Negro	1	0.66	0.79	0.3 – 1.61
Pt. scalare	1	6.3	5.77	2.74 – 9.94
MRCA (Peru & Manacapuru)	0.62	0.23	0.46	0.17 – 0.9
Manacapuru	1	0.21	0.23	0.07 – 0.51
Peru	1	0.20	0.23	0.07 – 0.51
MRCA (other)	0.39	0.26	0.46	0.17 – 0.9



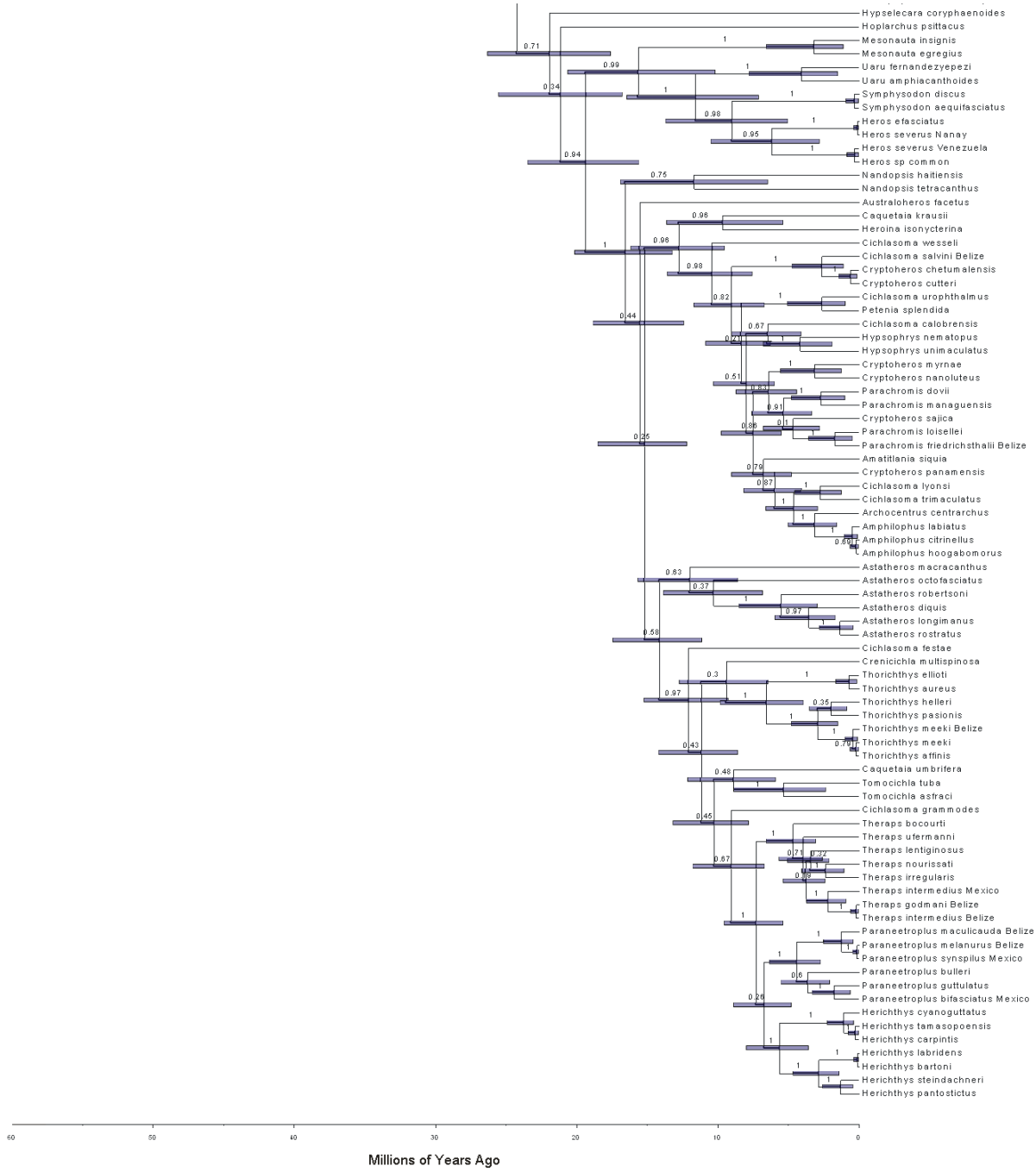


FIGURE 3.17 A time-calibrated maximum clade credibility Bayesian inferred phylogeny of the Neotropical subfamily Cichlinae. The phylogeny was based on the aligned mitochondrial Cyt *b* sequence matrix of 1,119 bp which consisted of 180 nucleotide sequences representative of seven outgroup taxa, 159 Neotropical reference taxa and 21 *Pterophyllum* taxa. Purple bars (95% HPD) represent the posterior distribution of divergence time estimates for their respective nodes. Bayesian posterior probabilities (BPP) ranging from 0.01-1.00 are indicated above branches. The taxonomic position of the monophyletic genus *Pterophyllum* and its three lineages (*Pt. leopoldi*, *Pt. altum* and *Pt. scalare*) is marked with a dark red triangle. The time-calibrated tree is shown with ages given in millions of years.

3.4 DISCUSSION

Based on the character variability of the five-gene sequence matrix it was apparent that the mitochondrial sequence data was the most variable. Together, the three mitochondrial (Cyt *b*, 16S rRNA and ND4) marker genes contributed 1,461 of the total 2,326 variable characters of the five-gene sequence matrix, thus contributing 62.8% (almost two thirds) to the overall sequence variability. The ND4 gene was found to be the most variable of the three mitochondrial genes. The two nuclear (S7 and RAG2) genes could only account for 865 of the total 2,326 variable characters of the five-gene sequence matrix, thus contributing 37.2% of the total sequence variability. Within the genus *Pterophyllum* there was even less sequence variation and the proportion of mitochondrial to nuclear gene variation was even greater (90:10).

The individual gene trees that were retrieved reflect their sequence variability. When the three mitochondrial gene trees were compared, the branch lengths of the 16S rRNA gene, averaged across the whole tree, were shorter compared to the Cyt *b* and ND4 gene trees indicating that the mitochondrial Cyt *b* and ND4 genes accumulated nucleotide substitutions at a faster rate than the 16S rRNA gene. Within the context of cichlid phylogenetic studies, the use of the mitochondrial Cyt *b* and ND4 genes are recommended to resolve the evolutionary relationships of more recently diverged species at the genus level. However, the use of the 16S rRNA gene is recommended for the resolution of the evolutionary relationships of more distantly related taxa. In contrast to the three mitochondrial genes, the two nuclear markers RAG2 and S7 appeared to be much slower evolving loci. This was particularly evident from the ML analysis performed on the concatenated nuclear sequence matrix which retrieved unresolved *Pt. altum* and *Pt. scalare* clades as polytomies. Hence, based on their shorter branch lengths and the nuclear character variability assessments, both the S7 and RAG2 loci appear to be more suitable for use in cichlid phylogenetic studies aimed to resolve the deeper evolutionary nodes i.e. the phylogenetic relationships of distantly related taxa.

Based on the concatenated five-gene phylogeny recovered from the ML analysis, three well supported species clades (*Pt. leopoldi*, *Pt. altum* and *Pt. scalare*) were retrieved within the genus *Pterophyllum*. *Pterophyllum leopoldi* formed the basal lineage whilst *Pt. altum* and *Pt. scalare* were retrieved in a well supported (89%) derived sister association. Species variation within the *Pt. altum* lineage was evident as two well supported clades were recovered. One clade contained *Pt. altum* species from its type locality i.e. Rio Atabapo and two tributaries (Rio Inirída and Rio Ventuari) of the Upper Rio Orinoco whilst the other contained fishes from the upper Rio Negro collected in the vicinity of San Carlos and San Felipe. As such, this study was the first to have included *Pt. altum* from its type locality i.e. Rio Atabapo and to show that the fishes from the upper Rio Negro were indeed *Pt. altum*. Furthermore, although the *Pt. scalare* lineage was unresolved, three groups could be retrieved containing most *Pt. scalare* including those from Peru, as well as from Santa Isabel and from Lago Manacapuru. The latter two groups were retrieved in a moderately supported sister association.

Based on the tree topology recovered from the ML analysis performed on the concatenated and aligned five-gene (Cyt *b*, 16S rRNA, ND4, S7 and RAG2) sequence matrix, the closely related South American species *Hoplarchus psittacus* and *Hypselecara coryphaenoides* were recovered as the basal lineage of the tribe Heroini. Furthermore, the genus *Pterophyllum* was retrieved as a well supported (100%) monophyletic clade nested in between the basal lineage and a clade containing the Mesonautines and Central American Heroines. This result is in agreement with previously published topologies of the Cichlinae (Farias et al., 2000; López-Fernández et al., 2010; Smith et al., 2008).

The parsimony analysis performed on the concatenated and aligned five-gene sequence matrix, retrieved a topology of which the backbone is well resolved with most nodes retrieved with 100% bootstrap support values. However, the positions of many clades shifted when compared to the ML phylogeny. Of importance to this study is the fact that the genus *Pterophyllum* was now retrieved as the basal lineage of the tribe Heroini instead of the closely related South American genera *Hoplarchus* and *Hypselecara*. A reasonable explanation for the changes in phylogenetic positioning of most clades could possible be that there is homoplasy in the combined five gene phylogeny which is supported by the low CI and RI values as shown in **Table 3.7**. It is well known that large data sets, analysed with parsimony, suffer from long branch attraction which may be the case in this instance.

The age estimates for the four Cichlidae sub-families (Cichlinae, Etroplinae, Pseudocrenilabrinae and Ptychochrominae) used to perform a time-calibrated BEAST analysis on the aligned Cyt *b* sequence matrix were obtained from Friedman et al. (2013). Friedman et al. (2013) used a comprehensive dataset which included 10 protein-coding nuclear genes for 158 species of Percomorpha (89 cichlid species). López-Fernández et al. (2013) used a Cretaceous origin (~130 Ma ago) of the Cichlidae based on geophysical calibrations (separation of Madagascar and Africa) in combination with the same fossil records used by Friedman et al. (2013). The assumption of the geophysical calibration date is extremely uncertain. Recently McMahan et al. (2013) using different marker genes of the cichlid groupings as López-Fernández et al. (2013) with dates obtained from cichlid fossils only, dated the Cichlidae radiation at 64.9 Ma ago (57.3-76.0 Ma ago; 95% HPD). This is in agreement with the dates obtained for the Cichlidae radiation by Friedman et al. (2013). However, the secondary dates for the Cichlidae radiation generated by Friedman et al. (2013) were used to date the Cyt *b* tree because attempts to date the complete five-gene phylogeny were unsuccessful for technical reasons which were not clear. Time constraints furthermore restricted attempts to date the Cyt *b* gene using the three cichlid fossil dates used by McMahan et al. (2013).

The topology of the time-calibrated Cyt *b* gene tree retrieved from the BEAST analysis was for the most part in agreement with the ML tree topology recovered for the five-gene sequence matrix. The evolution of the three species of the genus *Pterophyllum* appears to be in accordance with the major geological events of the Neogene period that had shaped the modern Amazonian landscape (Albert and Reis, 2011; Hoorn and Wesselingh, 2010; Hoorn et al., 2010, 1995). The dates obtained for the

age of the tribe Heroini by López-Fernández et al. (2013), McMahan et al. (2013) and Friedman et al. (2013) were ~75, 40 and 15 Ma ago respectively. The date for the tribe Heroini obtained in this study was ~24.2 Ma ago falls within the limits obtained in the three previously mentioned studies. According to the BEAST temporal divergence estimates, the ancestor of the genus *Pterophyllum* evolved at approximately 24.2 Ma ago during the early Miocene epoch. This ancestor underwent a speciation event during the late Miocene to early Pliocene at approximately 7.8 Ma ago (95% HPD: 12.66 – 4.06 Ma) to give rise to the species *Pt. leopoldi* and the MRCA of the two species *Pt. altum* and *Pt. scalare*. The ancestor of the genus *Pterophyllum* was most probably present in the Pebas system which had a northward course to the Llanos basin during this time. The Llanos basin may have drained the precursor of the Essequibo River which could account for *Pt. leopoldi*'s presence in that drainage system. However, the upper Essequibo and the Rio Branco tributary of the Amazon river are also currently connected by a swamp region which would allow migration of *Pt. leopoldi* into the Essequibo River.

The divergence of *Pt. altum* and *Pt. scalare* (7 to 2.5 Ma ago) coincides with the timing of the uplift of the Vaupes Arch (7-11 Ma ago) (Hoorn and Wesselingh, 2010). The latter resulted in the separation of the paleo-Amazon-Orinoco River into the modern Amazon and Orinoco drainage basins which would have allowed *Pt. altum* and *Pt. scalare* to evolve through allopatric speciation. The presence of *Pt. altum* in the upper Rio Negro shows that these two drainages were reconnected via the Casiquiare canal. Winemiller and Willis (2011) have discussed the implications of this reconnection for cichlid genera such as *Cichla* and *Satanoperca*. It appears that *Satanoperca daemon* has migrated through the Casiquiare to speciate in the Orinoco river (Winemiller and Willis, 2011). Members of the genus *Cichla* have also migrated through the Casiquiare to extend their ranges and possibly give rise to hybrids (Willis et al., 2007). However, none of these studies has included a dated phylogeny from which the date of the possible reconnection of the Amazon and Orinoco drainages could be derived. This is therefore the first study to hypothesize a middle to late Pleistocene date (1.61 – 0.3 Ma) for the connection of the Casiquiare canal to the upper Rio Negro and thus the reconnection of the Amazon and Orinoco drainages.

In conclusion, this study is the first to establish the relationships of the three *Pterophyllum* species to each other and the position of the genus within the tribe Herioni. Furthermore, the evolution of the three species of the genus *Pterophyllum* was found to be in agreement with the changes in the Amazonian landscape during the Neogene period as well as the uplift of the Vaupes Arch which separated the paleo-Amazon-Orinoco River into two distinct drainages. In addition it shows phylogenetic evidence for the reconnection of the Orinoco and Amazon river via the Casiquiare Canal. However, the sequence variability in the five-gene sequence matrix is not sufficient to resolve relationships between populations within *Pterophyllum* species. For this reason, additional variable markers will have to be sought, and in view of the imbalance between the sequence variability of the

nuclear and mitochondrial genes, more variable nuclear markers will have to be found for these purposes.

4. AN INVESTIGATION INTO THE VARIABILITY OF ADDITIONAL NUCLEAR MARKERS AND INCREASED TAXON SAMPLING TO AID IN RESOLVING INTERSPECIES RELATIONSHIPS WITHIN THE GENUS *PTEROPHYLLUM*

4.1 INTRODUCTION

In the previous part of this study (Chapter 3) it was determined that the five-gene sequence matrix could not sufficiently resolve the interspecies relationships of the genus *Pterophyllum*. Character variability within the five-gene sequence matrix revealed that the nuclear loci (RAG2 and S7) could only account for 37.2% of the total number of variable characters. When the character variability in the reduced five-gene sequence matrix in the genus *Pterophyllum* was assessed, these two nuclear loci contributed a mere 9.8% of the total variable characters. Hence, nuclear character variability was too low which resulted in the five-gene sequence matrix being dominated by the mitochondrial data.

Accordingly, the first aim of this part of the study was to identify additional variable nuclear markers which could contribute significantly to the total variability and thus increase the resolution of taxa within the genus *Pterophyllum*. The second aim of this part of the study was to investigate the intraspecies variability within the genus *Pterophyllum*. Based on the outcomes of these two, the third and final aim was to answer the following questions: Was there variation between the *Pt. altum* species from three tributaries (Rio Atabapo, Rio Inirida and Rio Ventuari) of the Upper Rio Orinoco? Was there variation between the *Pt. scalare* variants from different localities such as Peru, Lago Manacapuru and Santa Isabel? Was there any molecular evidence to support the hypothesis that *Pt. scalare* species from Santa Isabel may represent hybrids between *Pt. altum* and *Pt. scalare* species in the upper Rio Negro? Such evidence can be found if potential hybrid taxa appear in conflicting positions in the phylogenies derived from nuclear versus mitochondrial genes.

In order to achieve the first aim 13 potential nuclear markers (TmoM27, X-src, Sd11, Sd15, Sd23, Gpd2, GnRH3-3, Ednrb1, Mitfb, Ib1, RpL8, MpCS and CteO12) were identified. The first seven were identified from literature related to phylogenetic studies of closely related genera of the genus *Pterophyllum*. The last six were chosen based on personal communication with Dr. Stuart C. Willis who has had success with these loci in phylogenetic studies of the Neotropical genus *Cichla*. The character variability of these nuclear gene regions was assessed by sequencing at least one *Pt. altum* and one *Pt. scalare* sample. Once identified, additional variability within the nuclear marker genes was investigated across a much wider spectrum of *Pterophyllum* samples. As the main focus was on the genus *Pterophyllym*, the aligned and concatenated sequence matrix was reduced to consist only of the nucleotide sequences of the three *Pterophyllum* species and seven closely related Neotropical taxa

Astronotus ocellatus, *Heros severus*, *Hoplarchus psittacus*, *Hypselecara coryphaenoides*, *Mesonauta insignis*, *Symphysodon aequifasciatus* and *Uaru fernandezyepezi*. In addition, the sequence matrix of the most variable marker gene i.e. ND4, was expanded by increasing the number of *Pterophyllum* samples from two to a maximum of six representatives per collection locality, where possible, to assess intraspecific variation. The sequence data, generated from these analyses, were then used in an attempt to establish the phylogenetic relationships of the taxa within the genus *Pterophyllum*.

4.2 MATERIALS AND METHODS

4.2.1 SAMPLE COLLECTION

In addition to the 27 tissue samples used in the previous part of this study (see **section 3.2.2**, **Table 3.1**), 24 new tissue samples were obtained from various collection localities to be included in the mitochondrial ND4 sequence matrix in order to investigated intraspecies relationships within the genus *Pterophyllum* (**Table 4.1**). The relevant localities are shown on a map in **Figure 3.1**. The number of *Pt. leopoldi* samples was increased from two to four representatives. The *Pt. altum* samples were subdivided into two main groups: *Pt. altum* from the Upper Rio Orinoco (Rio Atabapo, Rio Inírida and Rio Ventuari) and the Upper Rio Negro (San Felipe). The *Pt. altum* representatives from the Upper Rio Orinoco were increased from eight to 19 whilst the *Pt. altum* representatives from the Upper Rio Negro were increased from two to six. In total, the *Pt. altum* representatives thus increased from 10 to 25. *Pterophyllum scalare* samples were representative of six localities namely Manaus, Guyana, Santa Isabel (Brazil), Lago Manacapuru (Brazil), Rio Xingu (Brazil) and Peru. The total number of *Pt. scalare* representatives could only be increased by four samples as access to additional samples was limited. In addition, the three outgroup species *Astronotus ocellatus*, *Heros severus* and *Symphysodon aequifasciatus* were sourced from local pet shops in South Africa.

4.2.2 DNA EXTRACTION

DNA from some of the 24 additional *Pterophyllum* tissue samples (**Table 4.1**) was extracted using a slightly modified Phenol-Chloroform-Isoamylalcohol protocol (Dorn et al., 2011). The reader is referred to **section 3.2.2** for a detailed description of the PCI methodology for DNA extraction. However, DNA yields following the PCI methodology were found to be low in some instances and as a result the DNeasy Blood and Tissue kit (Qiagen) was used for all further DNA extractions according to the manufacture's protocol for animal tissue (**Table 4.1**). The DNA concentration of all the samples was measured using the NanoDrop® (Inqaba Biotec) apparatus (see **section 3.2.2**).

Table 4.1: A list of the 51 tissue samples used in this part of the study. Information summarised includes the source (river system), method of DNA extraction, the resulting DNA concentrations and molecular markers analysed per sample. The additional 24 *Pterophyllum* samples included in this part of the study are highlighted in pink.

Sample nr	Species names	Source	Method of extraction*	[DNA] (µg/µL)	Molecular Markers Analysed:							
					16S rRNA	Cyt *b*	ND4	S7	RAG2	Mitfb	RpL8	ND4 expanded
	Pt. leopoldi											
1	*Pt. leopoldi* A	Aquarium Trade	PCI	8.0								X
2	*Pt. leopoldi* B	Aquarium Trade	PCI	10.3								X
3	*Pt. leopoldi* C	Aquarium Trade	DNeasy	-	X	X	X	X	X	X	X	X
4	*Pt. leopoldi* D	Aquarium Trade	DNeasy	-	X	X	X	X	X	X	X	X
	Pt. altum (Orinoco)											
5	*Pt. altum* Rio Atabapo 1	Rio Atabapo	DNeasy	5.9	X	X	X	X	X	X	X	X
6	*Pt. altum* Rio Atabapo 2	Rio Atabapo	DNeasy	5.6	X	X	X	X	X	X	X	X
7	*Pt. altum* Rio Atabapo 3	Rio Atabapo	DNeasy	2.7								X
8	*Pt. altum* Rio Atabapo 4	Rio Atabapo	DNeasy	49.9								X
9	*Pt. altum* Rio Atabapo 5	Rio Atabapo	DNeasy	23.1								X
10	*Pt. altum* Rio Inirida 1	Rio Inirida	DNeasy	-0.1		X	X		X			X
11	*Pt. altum* Rio Inirida 2	Rio Inirida	DNeasy	18.3		X	X	X	X			X
12	*Pt. altum* Rio Inirida 3	Rio Inirida	PCI	5.5								X
13	*Pt. altum* Rio Inirida 4	Rio Inirida	PCI	3.5								X
14	*Pt. altum* Rio Inirida 5	Rio Inirida	PCI	60.0	X							
15	*Pt. altum* Rio Inirida 6	Rio Inirida	PCI	89.4	X							
16	*Pt. altum* Rio Inirida 7	Rio Inirida	DNeasy	41.1								X
17	*Pt. altum* Rio Inirida 8	Rio Inirida	DNeasy	26.8								X
18	*Pt. altum* Rio Ventuari 2	Rio Ventuari	DNeasy	3.2								X
19	*Pt. altum* Rio Ventuari 3	Rio Ventuari	PCI	9.3	X	X	X	X	X			X
20	*Pt. altum* Rio Ventuari 4	Rio Ventuari	PCI	8.0	X	X	X	X	X			X
21	*Pt. altum* Rio Ventuari 7	Rio Ventuari	DNeasy	77.8						X	X	X
22	*Pt. altum* Rio Ventuari 8	Rio Ventuari	DNeasy	61.8						X	X	X
23	*Pt. altum* Rio Ventuari 10	Rio Ventuari	DNeasy	56.1								X
	Pt. altum (Negro)											
24	*Pt. altum* Rio Negro 1	Rio Negro	DNeasy	44.2	X	X	X	X	X	X	X	X
25	*Pt. altum* Rio Negro 2	Rio Negro	DNeasy	67.5	X	X	X	X	X	X	X	X

Sample nr	Species names	Source	Method of extraction*	[DNA] (µg/µL)	Molecular Markers Analysed:							
					16S rRNA	Cyt *b*	ND4	S7	RAG2	Mitfb	RpL8	ND4 expanded
26	*Pt. altum* Rio Negro 3	Rio Negro	DNeasy	92.7								X
27	*Pt. altum* Rio Negro 4	Rio Negro	DNeasy	34.7								X
28	*Pt. altum* Rio Negro 5	Rio Negro	DNeasy	92.6								X
29	*Pt. altum* Rio Negro 6	Rio Negro	DNeasy	130.1								X
	Pt. scalare											
30	*Pt. scalare* Germany B	Aquarium Trade	PCI	6.5	X	X	X	X	X			X
31	*Pt. scalare* Germany C	Aquarium Trade	PCI	10.4			X	X	X			X
32	*Pt. scalare* Germany D	Aquarium Trade	DNeasy	-	X	X				X	X	X
33	*Pt. scalare* Germany E	Aquarium Trade	DNeasy	-						X	X	X
34	*Pt. scalare* Germany F	Aquarium Trade	DNeasy	-								X
35	*Pt. scalare* Germany G	Aquarium Trade	DNeasy	-								X
36	*Pt. scalare* Guyana A	Aquarium Trade	PCI	7.1	X	X	X	X	X			X
37	*Pt. scalare* Guyana B	Aquarium Trade	PCI	6.1	X	X	X	X	X			X
38	*Pt. scalare* Santa Isabel A	Aquarium Trade	PCI	17.2	X			X	X			
39	*Pt. scalare* Santa Isabel B	Aquarium Trade	PCI	14.7	X		X	X	X			X
40	*Pt. scalare* Manacupuru A	Aquarium Trade	DNeasy	-	X	X	X	X	X	X	X	X
41	*Pt. scalare* Manacupuru B	Aquarium Trade	DNeasy	-	X	X	X	X	X	X	X	X
42	*Pt. scalare* Peru 1A	Aquarium Trade	PCI	15.1	X		X	X	X			X
43	*Pt. scalare* Peru 1B	Aquarium Trade	PCI	8.4	X							
44	*Pt. scalare* Peru 2A	Aquarium Trade	PCI	4.9		X	X	X	X			X
45	*Pt. scalare* Peru 2B	Aquarium Trade	PCI	6.4		X						
46	*Pt. scalare* Xingu A	Aquarium Trade	PCI	8.5								X
47	*Pt. scalare* Xingu B	Aquarium Trade	DNeasy	-	X	X	X	X	X			X
48	*Pt. scalare* Xingu C	Aquarium Trade	DNeasy	-	X	X	X	X	X			X
	Neotropical cichlids											
49	*Astronotus ocellatus*	Aquarium Trade	DNeasy	28.2						X	X	
50	*Heros severus*	Aquarium Trade	DNeasy	49.7						X	X	
51	*Symphysodon aequifasciatus* A	Aquarium Trade	DNeasy	13.0						X	X	

***DNeasy:** DNeasy Blood and Tissue kit**; PCI:** Phenol-Chloroform-Isoamylalcohol.

4.2.3 PRIMER SEQUENCES FOR MARKER GENE AMPLIFICATION

All primer pairs utilised for the amplification of the 13 nuclear markers (TmoM27, X-src, Sd11, Sd15, Sd23, Gpd2, GnRH3-3, Ednrb1, Mitfb, Ib1, RpL8, MpCS and CteO12) investigated in this study is summarised in **Table 4.2**. The reader is referred to **section 3.2.4** and **Table 3.2** for a detailed summary of the original five molecular markers (Cyt *b*, 16S rRNA, ND4, RAG2 and S7) additionally used in this study. All primers pairs were synthesized by the DNA Synthesis Laboratory, Department of Molecular and Cell Biology, University of Cape Town, South Africa. Stock solutions of 100µM were prepared for each of the primers by calculating their extinction coefficients at 260nm according to the following equation:

$$C_{primer} = A_{260} * \frac{100\ \mu M}{1.5(A)+0.71(C)+1.2(G)+0.84(T)}$$

Accordingly, working solutions of 20 µM were prepared from the stock solutions. All primers, stock solutions and working solutions were stored at -20°C until further use.

4.2.4 AMPLIFICATION OF THE 13 NUCLEAR MARKERS

Amplification of the 13 nuclear markers was carried out individually using the Multiplex PCR Kit (Qiagen®) with their respective primer pairs as specified in **Table 4.2**. In short, each PCR mixture contained the following: 2.2 µL RNase-free water, 1.5 µL (5X) Q-Solution, 1.3 µL (20 µM) forward primer, 1.3 µL (20 µM) reverse primer and 7.7 µL (2X) Qiagen® Multiplex PCR Master Mix. To this mixture, 1 µL of undiluted extracted DNA was added to obtain a final PCR reaction volume of 15 µL. Amplification was carried out according to the manufacturer's protocol consisting of a pre-denaturation step at 95°C for 15 min followed by 40 cycles of DNA amplification at 94°C for 30 seconds, 46°C - 56°C (Table 8) for 90 seconds and 72°C for 70-90 seconds, depending on the target sequence's length. A final extension completion period followed at 72°C for 10 minutes. Where necessary, samples were held at 15°C for no longer than one hour following completion of amplification. Samples were then stored at 4°C until further use. Amplification was performed using either a Veriti™ 96-Well Thermal Cycler (Applied Biosystems, Pty. Ltd) or a 2720™ Thermal Cycler (Applied Biosystems, Pty. Ltd). Individual gene amplifications were all optimised for optimal annealing temperatures in order to eliminate multiple band formation. Negative controls were run during these optimisation processes and consistently lead to no amplification. Amplification product identity was confirmed by sequencing which is described hereafter as no positive controls were available from other sources.

4.2.5 AMPLIFICATION OF THE MITOCHONDRIAL ND4 GENE

The mitochondrial ND4 gene was amplified from 43 taxa in total (**Table 4.1**). Twenty-two of these taxa were investigated in the previous study (see **section 3.2.2**, **Table 3.1**) whilst in this part of the

study the ND4 sequence matrix was expanded to include an additional 21 taxa in order to investigate intraspecies variability. Amplification proceeded according to the guidelines as stipulated for the amplification of the 13 nuclear markers as described above. The reader is referred to **section 3.2.4** and **Table 3.2**, in this thesis, for a summary of the primer sequences used to amplify the mitochondrial ND4 gene and its expected amplicon size.

4.2.6 Electrophoresis And PCR Product Purification

All PCR products were analysed on 1% agarose gels which are prepared as follows: 0.7 g of agarose powder was weighed off and added to 70 mL 1X casting buffer. This mixture was then gently swirled to obtain a homogenized solution which was then heated for two minutes at medium power in a microwave oven. The mixture was then allowed to cool to approximately 60°C (15 min) after which 7 µL (10,000X) GelRed™ Nucleic Acid Gel Stain (Biotium) was added. The liquid was then poured into a casting plate, the appropriate plastic comb/spacer inserted and a time period of 20-30 min allowed for the gel to solidify. The gel slab was then immersed in ±600 mL of 1X running buffer after which a sufficient volume of PCR product was then mixed with 1 µL loading buffer and loaded onto the 1% agarose gel. Electrophoresis was then performed at 110 V for 65 min using an EPS 601 (Amersham Pharmacia Biotech, Inc.) electrophoresis setup. Visualisation of the PCR amplicons was performed using an E-Box CN-1000 UV transilluminator and portable darkroom (Vilber Lourmat). The subsequent purification of the PCR amplicons was accomplished by means of the Wizard® SV Gel and PCR Clean-Up System (Promega) according to the manufacturer's instructions. Where required, the purified PCR amplicons were concentrated using a SpeedVac Concentrator (Sarvant Instruments, Farmingdale, N.Y.). Once concentrated, 2 µL of each sample was analysed on a 1% agarose gel to determine DNA concentration. A detailed summary of the constituents of each of the buffers used during electrophoresis and PCR product purification are presented in **Addendum A**.

4.2.7 Cycle Sequencing Of Purified PCR Fragments

Selected nuclear markers were bi-directionally sequenced using the primers as summarised in **Table 4.2**. The reader is referred to **section 3.2.6** for a detailed decription of the cycle sequencing methodology used in this study.

Table 4.2: A list of the 13 nuclear markers investigated in this study. Summarised are the primer names, primer sequences, primer lengths, literature sources, annealing temperatures, amplicon lengths and aligned lengths.

Molecular Marker	Locus	Primer name	Primer sequence (5' - 3')	Length (bp)	Reference	Annealing temperature (°C)	Amplicon length (bp)	Aligned Length (bp)
TmoM27 microsatellite flanking region	TmoM27	TmoM27_f1	AGG CAG GCA ATT ACC TTG ATG TT	23	(Zardoya et al., 1996b)	52	350	333
		TmoM27_r1	TAC TAA CTC TGA AAG AAC CTG TGA T	25				
X-src	X-src	X-src D cich	ATG TCC CCT GAG GCT TTC CT	20	(Sides and Lydeard, 2000)	*ND	650	-
		X-src C mod	CTC AAT CAG GCG AGC CAA ACC AAA ATC	27				
Symphysodon discus microsatellite 11	Sd11	Sd11_F	GAC AGC TGC AGA CAG TCT TTT T	22	(Amado et al., 2008)	*ND	180	-
		Sd11_R	CCA ATC TCA TTG TAC ACC TCC A	22				
Symphysodon discus microsatellite 15	Sd15	Sd15_F	TCC CGA GAT GTT TAA TGC TG	20	(Amado et al., 2008)	*ND	160-210	-
		Sd15_R	CAA ACA TTT CCT GAA ATT CAA CC	23				
Symphysodon discus microsatellite 23	Sd23	Sd23_F	CAG ACT CCA AGC TGT GCT TT	20	(Amado et al., 2008)	55	350	307
		Sd23_R	GAC TTG CTA GCA GCT CAG GAC	21				
Glyceraldehyde 3-phosphate dehydrogenase 2	Gpd2	Gpd2_F	GCC ATC AAT GAC CCC TTC ATC G	22	(Hassan et al., 2002)	55	350	-
		Gpd2_R	TTG ACC TCA CCC TTG AAG CGG CCG	24				
Gonadotropin-releasing hormone 3	GnRH3-3	GnRH3_F	GCC CAA ACC CAA GAG AGA CTT AGA CC	26	(Hassan et al., 2002)	55	350	365
		GnRH3_R	TTC GGT CAA ATG ACT GGA ATC ATC	24				
Endothelin receptor b1	Ednrb1	ROSE_F1	CAG GGA TCG GAG TTC CAA AAT	21	(Won et al., 2006)	53	850	763
		ROSE_R1	AAC GGC AGG CAG AAA TAC AC	20				

Molecular marker	Locus	Primer name	Primer sequence (5' – 3')	Length (bp)	Reference	Annealing temperature (°C)	Amplicon length (bp)	Aligned Length (bp)
Microphthalamia b	mitfb	Mitfb_f1	CAG CCC TAT GGC CTT ATT GA	20	(Won et al., 2006)	52	650	698
		Mitfb_r1	CCT TTT GAT GTT TGG CAG GT	20				
Karyopherin (importin) beta 1	8680e3	Ib1_F	GGA GGA GAR TTY AAG AAG TAY CTG GAC AT	29	(Li et al., 2010)	58	450	432
		Ib1_R	CSC CCT TCA GGC CCT GGA TGA T	22				
60S ribosomal protein L8	14867e1	RpL8_F	CCA CAA RTA CAA GGC CAA GAG RAA CTG	27	(Li et al., 2010)	58	800	861
		RpL8_R	GTT CTC CTT STC CTG SAC GGT CTT	24				
Microprocessor complex subunit DGCR8-like	35564e5	MpCS_F	AAG ACT CAA GRG TGT AYG AGC TGA CCA A	28	(Li et al., 2010)	58	650	630
		MpCS_R	CAT GTC ATC ACR TAT TCR CTC TTC TGR TT	29				
Cichla temensis clone O12	CteO12	Cte012_F	TGG ATG CAG TAT GCA AAC ACT	21	(Willis, 2011)	*ND	687	-
		Cte012_R	AGA GGT GAC CAC GCT TTG AG	20				

*ND: Not determined

4.2.8 Nucleotide Sequence Analysis And Alignment Of The Seven Molecular Markers

All subsequent chromatograms were edited using ChromasPro (v1.5, Technelysium, Pty. Ltd.) and the nucleotide sequences exported as .txt files for analysis in BioEdit® Sequence Alignment Editor (v7.1.3.0) (Hall, 1999). Seven alignment matrices representing each of the seven molecular markers: Cyt *b*, 16S rRNA, ND4, S7, RAG2, and the two newly identified variable nuclear markers Mitfb and RpL8 were generated. Seven Neotropical cichlids (*Astronotus ocellatus*, *Heros severus*, *Hoplarchus psittacus*, *Hypselecara coryphanoides*, *Mesonauta insignis*, *Symphysodon aequifasciatus* and *Uaru fernandezyepezi*) were selected as the outgroup species. Nucleotide sequences of the Mitfb and RpL8 loci were not available for the seven outgroup species on Genbank. The Mitfb and RpL8 loci were therefore amplified from *Astronotus ocellatus*, *Heros severus* and *Symphysodon aequifasciatus* and sequenced to be included in their respective matrices as outgroup species (**Table 4.1**). As material of all of the other species of which López-Fernández et al. (2010) was not available the outgroups were limited to the seven aforementioned species.

In addition, two representatives from each of the six sources (Manaus, Guyana, Santa Isabel (Brazil), Lago Manacapuru (Brazil), Rio Xingu (Brazil) and Peru, see **section 4.2.1**) were chosen to represent the three species of the genus *Pterophyllum* (**Table 4.1**). Subsequently, the generated nucleotide sequences together with the reference sequences of Neotropical outgroup cichlids were aligned in their respective matrices using the alignment tool Clustal W (v1.4) (Thompson et al., 1994) in BioEdit® Sequence Alignment Editor (v7.1.3.0) (Hall, 1999). Finer alignment of the seven separate matrices was done by eye. The Genbank accession numbers of the seven outgroup species for each of the five molecular markers (Cyt *b*, 16S rRNA, ND4, S7 and RAG2) are summarised in **Addendum B** (López-Fernández et al., 2010).

4.2.9 Nucleotide Sequence Analysis And Alignment Of The Expanded Mitochondrial ND4 Data Matrix

As this study involved expanding the ND4 sequence matrix from two to six representatives per source (Manaus, Guyana, Santa Isabel (Brazil), Lago Manacapuru (Brazil), Rio Xingu (Brazil) and Peru, see **section 4.2.1**) where possible, all the ND4 nucleotide sequences used in the previous study (Chapter 3) were included in this study. Furthermore, the chromatograms of all the newly synthesised ND4 nucleotide sequences were edited using ChromasPro (v1.5, Technelysium, Pty. Ltd.) and exported as .txt files for further analysis. The seven Neotropical cichlids (*Astronotus ocellatus*, *Heros severus*, *Hoplarchus psittacus*, *Hypselecara coryphanoides*, *Mesonauta insignis*, *Symphysodon aequifasciatus* and *Uaru fernandezyepezi*) were selected as the preferred outgroups for the expanded ND4 sequence matrix. The outgroup species together with the .txt files of the previously used and newly synthesised ND4 nucleotide sequences were imported into the BioEdit® Sequence Alignment Editor (v7.1.3.0)

(Hall, 1999) and aligned using the alignment tool Clustal W (v1.4) (Thompson et al., 1994). Finer alignment of the expanded ND4 sequence matrix was done by eye. The Genbank accession numbers for the ND4 gene of each of the seven outgroup species are listed in **Addendum B** (López-Fernández et al., 2010).

4.2.10 PHYLOGENETIC ANALYSIS OF THE SEVEN ALIGNMENT MATRICES

4.2.10.1 Character variability assessment

Character variability was assessed using the software package PAUP* (v4.0b10) (Swofford, 2003) and SequenceMatrix (Vaidya et al., 2011). The seven aligned matrices were exported from BioEdit® Sequence Alignment Editor (v7.1.3.0) (Hall, 1999) as Nexus files. First, each of the seven reduced and aligned matrices were analysed separately. This was followed by character variability analyses performed on two concatenated sequence matrices. The first sequence matrix contained the reduced, concatenated and aligned mitochondrial (Cyt *b*, 16S rRNA and ND4) sequence matrices whilst the second consisted of the reduced, concatenated and aligned nuclear (RAG2, S7, Mitfb and RpL8) sequence matrices. Subsequently, the character variability of the five- and seven-gene sequence matrices were summarised to assess the change in overall character variability due to the inclusion of the two additional nuclear markers Mitfb and RpL8.

4.2.10.2 Maximum likelihood (ML) analysis

For maximum likelihood analyses each of the seven aligned sequence matrices were trimmed at both the 5’ and 3’ ends so as to exclude missing characters. The seven trimmed sequence matrices were then concatenated according to the same methodology applied for the character variability assessment. Maximum likelihood analysis was performed using RAxML-HPC2 on XSEDE version 8.0.24 (Stamatakis, 2014) available on The CIPRES Science Gateway V.3.3 (Miller et al., 2010). The species *Astronotus ocellatus* was chosen as the preferred outgroup. Each of the three focal datasets was partitioned using a mixed/partitioned model specifying each of the seven sequence matrices as independent partitions.

4.2.10.3 Generation of phylogenetic trees

The final trees recovered from each ML analysis performed were displayed and analysed using the online tool iTOL (interactive tree of life) v3.0 (Letunic and Bork, 2011, 2007). In setting the basic controls, the normal mode was selected and it was specified to not align leaf labels. The advanced control settings involved specifying all bootstrap values >50% to be displayed as text above each branch. Under the display controls the font size factor was kept at its default value of 1.0 X and the tree scale factor changed to 0.7 X.

4.2.11 Phylogenetic Analysis Of The Expanded Mitochondrial ND4 Alignment Matrix

Maximum likelihood analyses were performed on the expanded mitochondrial ND4 alignment matrix following the same methodology as described for the phylogenetic analysis of the seven alignment matrices above.

4.3 RESULTS

4.3.1 DNA Extraction

DNA was successfully isolated from 41 of the 51 tissue samples, excluding the 10 lyophilised samples received from Dr. Dirk Neumann. The DNA concentrations of the 41 samples are presented in **Table 4.1**. Compared to the PCI-method, the DNeasy Blood and Tissue kit gave an average DNA yield of 42.0 µg/µL DNA per sample whereas the PCI-method gave an average DNA yield of 15.8 µg/µL DNA per sample. This was almost three times less than that of the DNeasy Blood and Tissue which, in addition, required less starting material than the PCI-method.

4.3.2 Amplification And Gel Electrophoresis Of The 13 Nuclear Markers

Of the 13 primer pairs investigated **Sd15_F - Sd15_R** and **CteO12_F – CteO12_R** did not give amplification products. These results are shown in **Figures 4.1** and **4.2** respectively.

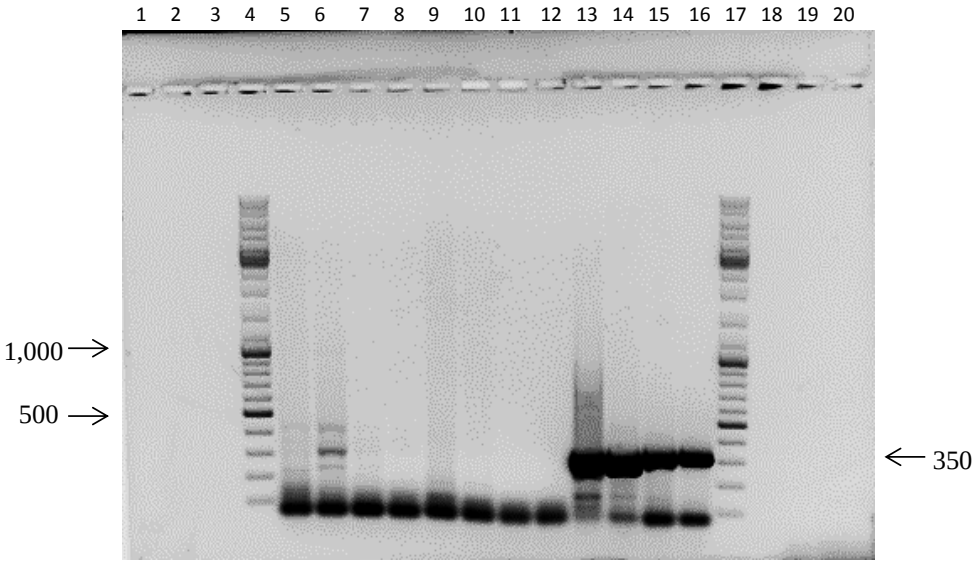


FIGURE 4.1 An image of a 1% agarose gel showing the separation of the PCR amplification products of the attempted amplifications of the three nuclear markers: Sd11, Sd15 and Sd23. The expected amplification product sizes were: Sd11, 180 bp; Sd15, 160-210 bp and Sd23, 350 bp. The four samples investigated for each of the three nuclear loci consisted of two *Pt. scalare* and two *Pt. altum* samples respectively. Lanes 1-3: open, lane 4: DNA ladder (GeneRuler™ DNA Ladder Mix), lanes 5-8: Sd11, lanes 9-12: Sd15, lanes 13-16: Sd23, lane 17: DNA ladder (GeneRuler™ DNA Ladder Mix) and lanes 18-20: open.

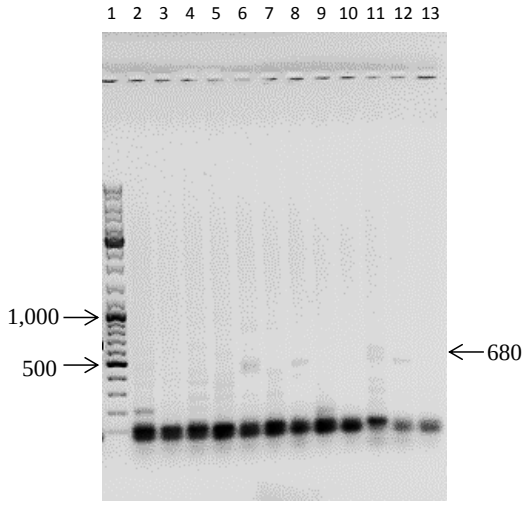


FIGURE 4.2 An image of a 1% agarose gel showing the separation of the PCR amplification products of the attempted amplification of the nuclear marker CteO12. Amplification was carried out along a temperature gradient of 53°C - 58°C with 1°C increments. The expected amplification product size was 680 bp. The two respective species *Pt. altum* and *Pt. scalare* were investigated at each temperature interval. Lane 1: DNA ladder (GeneRuler™ DNA Ladder Mix), lanes 2-3: CteO12 at 53°C, lanes 4-5: CteO12 at 54°C, lanes 6-7: CteO12 at 55°C, lanes 8-9: CteO12 at 56°C, lanes 10-11: CteO12 at 57°C and lanes 12-13: CteO12 at 58°C.

Furthermore, the three primer pairs **Gpd2_F – Gpd2_R**, **Sd11_F – Sd11_R** and **X-src D cich – X-src C mod** gave poor amplification. The amplification results for Sd11 are shown in **Figure 4.1** and that of Gpd2 and X-src are shown in **Figures 4.3** and **4.4** respectively. In addition to amplification being poor, it was also inconsistent as only one of the two samples investigated for each of the three nuclear markers (Gpd2; *Pt. altum*, Sd11; *Pt. scalare* and X-src; *Pt. scalare*) gave amplification products. For this reason the PCR amplicons were regarded not suitable for sequencing due to either a very low DNA concentration (Sd11 and Gpd2) or the presence of multiple bands (X-src).

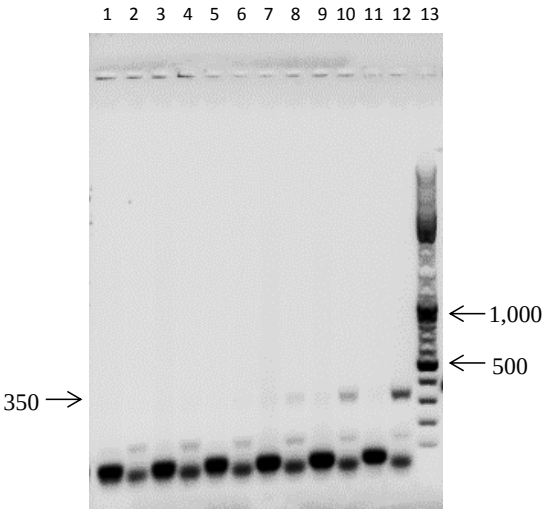


FIGURE 4.3 An image of a 1% agarose gel showing the separation of the PCR amplification products of the amplification of the nuclear marker Gpd2. Amplification was carried out along a temperature gradient of 50°C - 55°C with 1°C increments. The expected amplification product size was 350 bp.The two species investigated at each temperature interval were *Pt. scalare* and *Pt. altum* respectively. Lanes 1-2: Gpd2 at 50°C, lanes 3-4: Gpd2 at 51°C, lanes 5-6: Gpd2 at 52°C, lanes 7-8: Gpd2 at 53°C, lanes 9-10: Gpd2 at 54°C, lanes 11-12: Gpd2 at 55°C and lane 13: DNA ladder (GeneRuler™ DNA Ladder Mix).

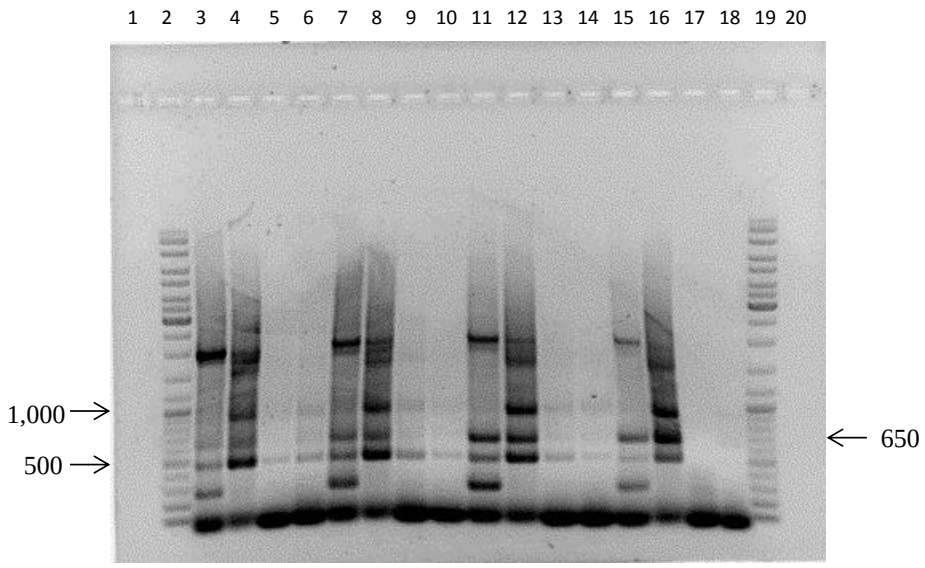


FIGURE 4.4 An image of a 1% agarose gel showing the separation of the PCR amplification products of the amplification of the nuclear marker X-src. Amplification was carried out along a temperature gradient of 46°C - 52°C with 2°C increments. The expected amplification product size was 650 bp. The four samples investigated at each of the temperature intervals consisted of two *Pt. scalare* and two *Pt. altum* samples respectively. Lane 1: open, lane 2: DNA ladder (GeneRuler™ DNA Ladder Mix), lanes 3-6: X-src at 46°C, lanes 7-10: X-src at 48 °C, lanes 11-14: X-src at 50°C, lanes 15-18: X-src at 52°C, lane 19: DNA ladder (GeneRuler™ DNA Ladder Mix) and lane 20: open.

Eight primer pairs (ROSE_F1-ROSE_R1, GnRH3_F-GnRH3_R, Ib1_F–Ib1_R, MpCS_F–MpCS_R, Sd23_F–Sd23_R, TmoM27_f1–TmoM27_r1, Mitfb_f1–Mitfb_r1 and RpL8_F–RpL8_R) gave sufficient amplification products. These are shown in **Figure 4.1**(Sd23) and **Figures 4.5** to **4.11**.

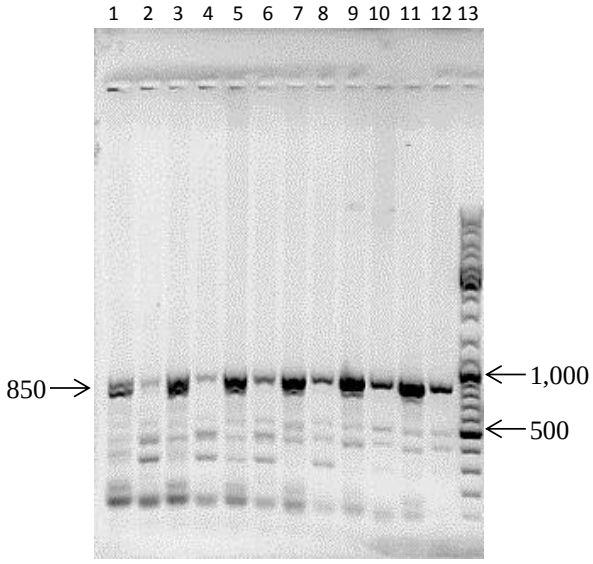


FIGURE 4.5 An image of a 1% agarose gel showing the separation of the PCR amplification products of the amplification of the nuclear marker Ednrb 1. Amplification was carried out along a temperature gradient of 48°C - 53°C with 1°C increments. The expected amplification product size was 850 bp. The two species investigated at each of the temperature intervals were *Pt. scalare* and *Pt. altum* respectively. Lanes 1-2: Ednrb 1 at 48°C, lanes 3-4: Ednrb 1 at 49°C, lanes 5-6: Ednrb 1 at 50°C, lanes 7-8: Ednrb 1 at 51°C, lanes 9-10: Ednrb 1 at 52°C, lanes 11-12: Ednrb 1 at 53°C and lane 13: DNA ladder (GeneRuler™ DNA Ladder Mix).

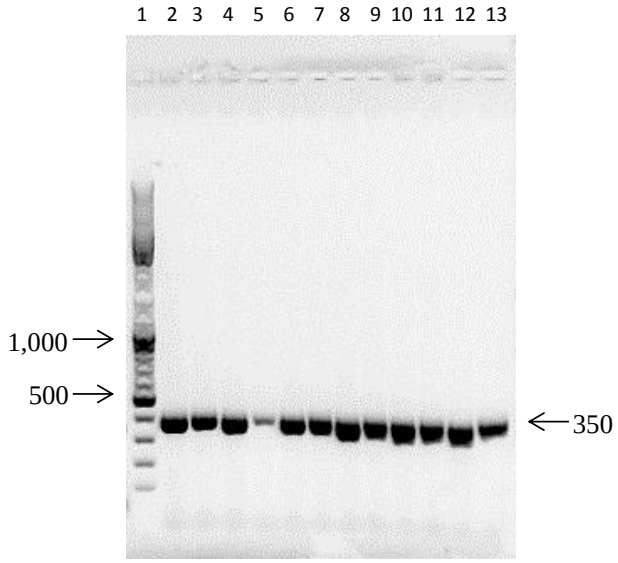


FIGURE 4.6 An image of a 1% agarose gel showing the separation of the PCR amplification products of the amplification of the nuclear marker GnRH3-3. Amplification was carried out along a temperature gradient of 51°C - 56°C with 1°C increments. The expected amplification product size was 350 bp. The two species investigated at each of the temperature intervals were *Pt. scalare* and *Pt. altum* respectively. Lane 1: DNA ladder (GeneRuler™ DNA Ladder Mix), lanes 2-3: GnRH3-3 at 51°C, lanes 4-5: GnRH3-3 at 52°C, lanes 6-7: GnRH3-3 at 53°C, lanes 8-9: GnRH3-3 at 54°C, lanes 10-11: GnRH3-3 at 55°C and lanes 12-13: GnRH3-3 at 56°C.

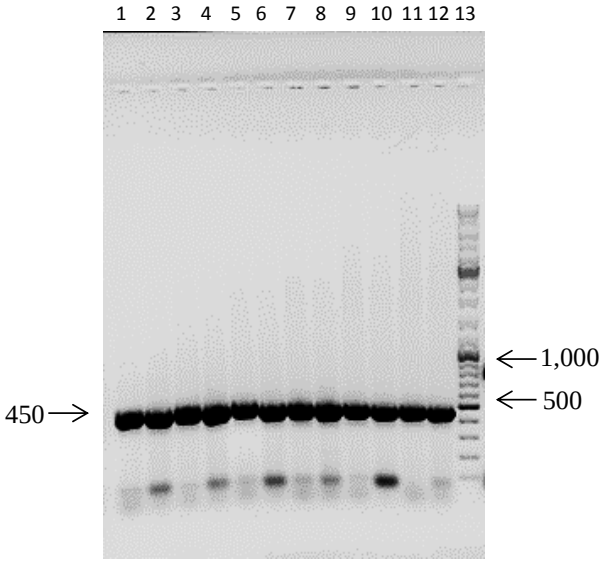


FIGURE 4.7 An image of a 1% agarose gel showing the separation of the amplification products of the amplification of the nuclear marker Ib1. Amplification was carried out along a temperature gradient of 53°C - 58°C with 1°C increments. The expected amplification product size was 450 bp. The two species investigated at each of the temperature intervals were *Pt. altum* and *Pt. scalare* respectively. Lanes 1-2: Ib1 at 53°C, lanes 3-4: Ib1 at 54°C, lanes 5-6: Ib1 at 55°C, lanes 7-8: Ib1 at 56°C, lanes 9-10: Ib1 at 57°C, lanes 11-12: Ib1 at 58°C and lane 13: DNA ladder (GeneRuler™ DNA Ladder Mix).

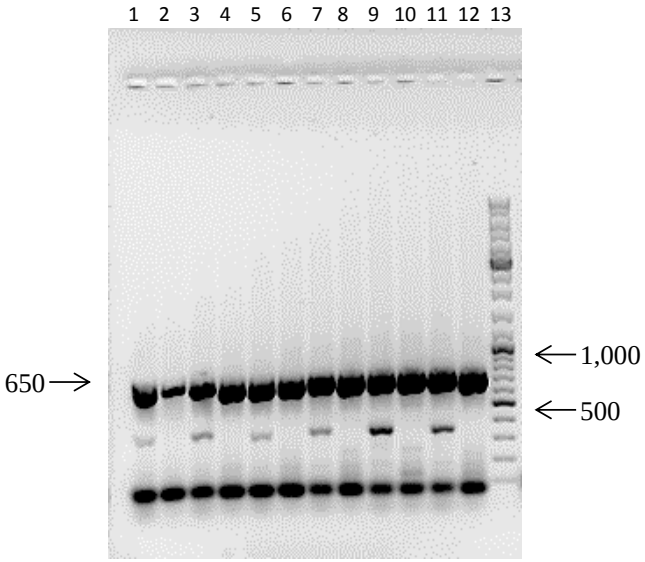


FIGURE 4.8 An image of a 1% agarose gel showing the separation of the amplification products of the amplification of the nuclear marker MpCS. Amplification was carried out along a temperature gradient of 53°C - 58°C with 1°C increments. The expected amplification product size was 650 bp. The two species investigated at each of the temperature intervals were *Pt. altum* and *Pt. scalare* respectively. Lanes 1-2: MpCS at 53°C, lanes 3-4: MpCS at 54°C, lanes 5-6: MpCS at 55°C, lanes 7-8: MpCS at 56°C, lanes 9-10: MpCS at 57°C, lanes 11-12: MpCS at 58°C and lane 13: DNA ladder (GeneRuler™ DNA Ladder Mix).

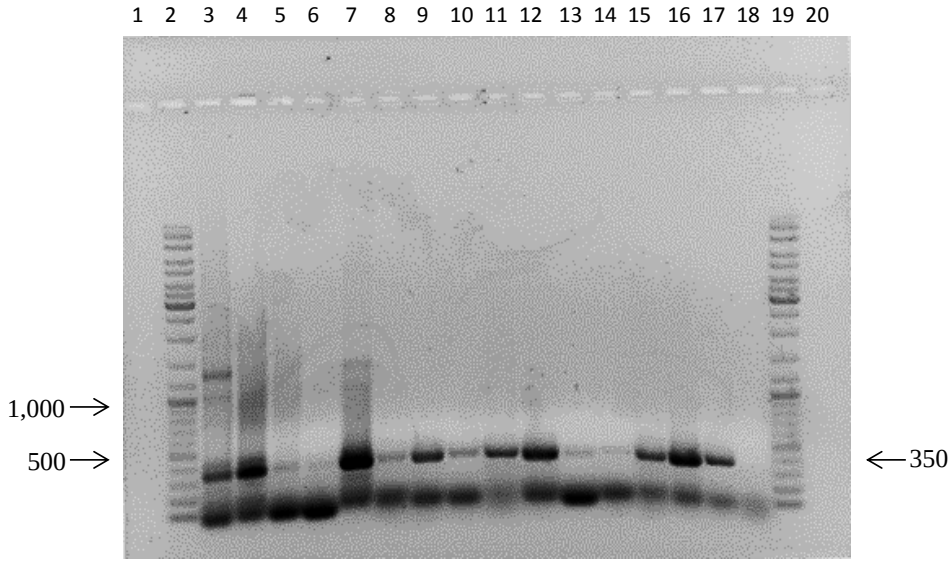


FIGURE 4.9 An image of a 1% agarose gel showing the separation of the amplification products of the amplification of the nuclear marker TmoM27. Amplification was carried out along a temperature gradient of 46°C - 52°C with 2°C increments. The expected amplification product size was 350 bp. The four samples investigated at each of the temperature intervals consisted of 2 *Pt. scalare* and 2 *Pt. Altum* samples respectively. Lane 1: open, lane 2: DNA ladder (GeneRuler™ DNA Ladder Mix), lanes 3-6: TmoM27 at 46°C, lanes 7-10: TmoM27 at 48 °C, lanes 11-14: TmoM27 at 50°C, lanes 15-18: TmoM27 at 52°C, lane 19: DNA ladder (GeneRuler™ DNA Ladder Mix) and lane 20: open.

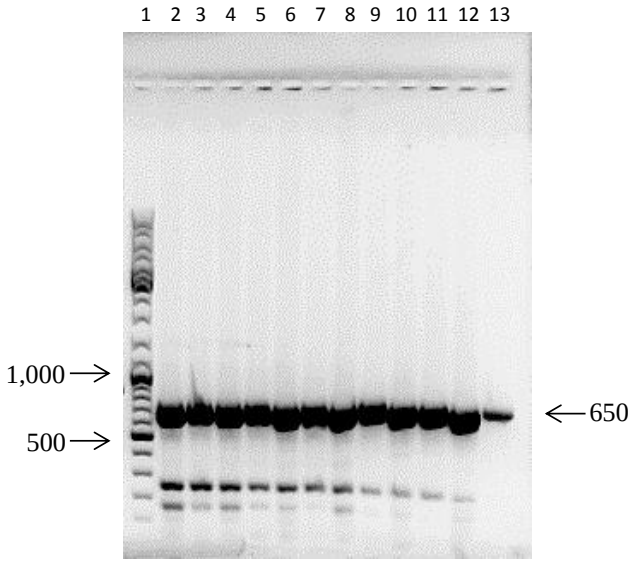


FIGURE 4.10 An image of a 1% agarose gel showing the separation of the amplification products of the amplification of the nuclear marker Mitfb. Amplification was carried out along a temperature gradient from 48°C - 53°C with 1°C increments. The expected amplification product size was 650 bp. The two species investigated at each of the temperature intervals were *Pt. scalare* and *Pt. altum* respectively. Lane 1: DNA ladder (GeneRuler™ DNA Ladder Mix), lanes 2-3: Mitfb at 48°C, lanes 4-5: Mitfb at 49°C, lanes 6-7: Mitfb at 50°C, lanes 8-9: Mitfb at 51°C, lanes 10-11: Mitfb at 52°C and lanes 12-13: Mitfb at 53°C.

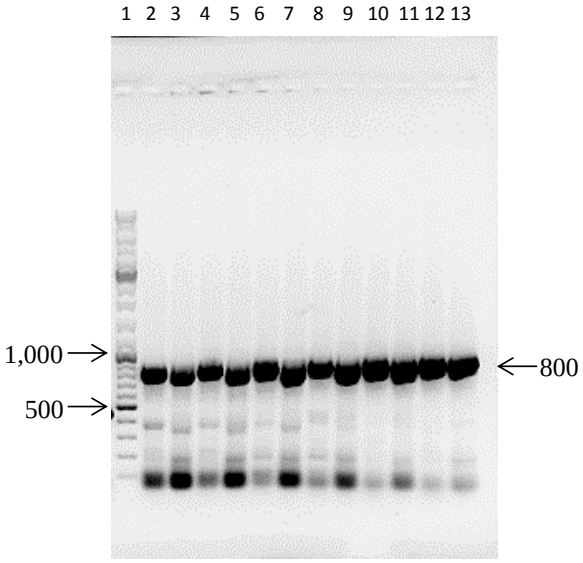


FIGURE 4.11 An image of a 1% agarose gel showing the separation of the amplification products of the amplification of the nuclear marker RpL8. Amplification was carried out along a temperature gradient from 53°C - 58°C with 1°C increments. The expected amplification product size was 800 bp. The two species investigated at each of the temperature intervals were *Pt. altum* and *Pt. scalare* respectively. Lane 1: DNA ladder (GeneRuler™ DNA Ladder Mix), lanes 2-3: RpL8 at 53°C, lanes 4-5: RpL8 at 54°C, lanes 6-7: RpL8 at 55°C, lanes 8-9: RpL8 at 56°C, lanes 10-11: RpL8 at 57°C and lanes 12-13: RpL8 at 58°C.

As can be seen from their respective gel images amplification products of each of the seven nuclear loci were obtained. Furthermore, the amplification of each of the seven nuclear markers (Sd23, Ednrb 1, GnRh3-3, Ib1, MpCS, Mitfb and RpL8) was consistent across their respective temperature gradient intervals. The only exception was the amplification of Tmo-M27. Although its amplification products were of the expected amplicon size (350 bp), amplification was inconsistent across the chosen

temperature gradient. The optimum annealing temperatures (giving one clear band) for each of the eight nuclear loci are presented in **Table 4.2**. A slight degree of non-specific amplification was observed at the optimum annealing temperatures of the four nuclear markers Ednrb1, MpCS, Mitfb and RpL8. However, the DNA concentrations of these non-specific amplicons were regarded to be insignificant when compared to that of the PCR amplicons of expected size.

4.3.3 Amplification And Gel Electrophoresis Of The Mitochondrial ND4 Gene

The reader is referred to **Figure 3.4** (Chapter 3) for a visual representation of the separation of the amplification products of the partial amplification of the mitochondrial ND4 gene on a 1 % agarose gel.

4.3.4 Nucleotide Sequence Analysis And Alignment Of The Seven Sequence Matrices

4.3.4.1 Identification of the most variable nuclear markers

For each of the eight nuclear loci (Ednrb 1, GnRH3-3, Ib1, MpCS, Sd23, TmoM27, Mitfb and RpL8) for which suitable PCR amplification products were generated, one specimen of *Pt. altum* and *Pt. scalare* each was used to investigate the sequence variability of each gene. The resulting DNA sequences for both *Pt. altum* and *Pt. scalare* were examined to give an indication of how much variability was present in each of the eight nuclear markers. **Table 4.3** summarises the total number of variable characters across the aligned sequence length of each nuclear loci.

Table 4.3 A summary of the total number of variable sites observed across the aligned length of each of the eight nuclear markers investigated. Summarised are the loci names, number of variable sites, aligned sequence lengths (in base pairs) and percentage variability.

Loci	Number of variable sites	Aligned length (bp)	% Variability
Ednrb 1	4	732	0.55
GnRH3-3	3	364	0.82
Ib1	3	434	0.69
MpCS	4	630	0.63
Sd23	0	306	-
TmoM27	0	312	-
Mitfb	6	698	0.86
RpL8	8	850	0.94

Accordingly, the two nuclear markers Mitfb and RpL8 were identified as the most variable of the eight nuclear markers investigated. The remaining six nuclear markers Ednrb 1, GnRH3-3, Ib1, MpCS,

Sd23 and TmoM27, of which Ednrb 1 was the least variable, showed inadequate variability and therefore were not included in the subsequent phylogenetic analysis.

4.3.4.2 Mitochondrial matrices

The aligned Cyt *b* sequence matrix of 1121 bp consisted of 28 nucleotide sequences representative of the seven Neotropical outgroup taxa and 21 *Pterophyllum* taxa (**Addendum D**). Optimal alignment of the 28 nucleotide sequences was obtained without introducing gaps into the sequence matrix and the outgroup nucleotide sequences contained only a few ambiguous base pairs. The Cyt *b* gene appeared to be a very fast evolving gene as the aligned sequence matrix contained numerous variable characters across its length. From the latter the nucleotide differences between the three *Pterophyllum* species and the outgroup taxa were clearly visible. The aligned 16S rRNA sequence matrix of 562 bp consisted of 30 nucleotide sequences representative of the seven Neotropical outgroup taxa and 23 *Pterophyllum* taxa (**Addendum E**). Only a few gaps (less than 5% of the aligned sequence length) were introduced to obtain optimal alignment of the 30 nucleotide sequences. Compared to the Cyt *b* and ND4 sequence matrices, the 16S rRNA gene had fewer variable sites across its length and thus, appeared to be the slowest evolving mitochondrial gene. In addition, the majority of variable characters observed across its length appeared to be shared between the outgroup taxa and the three *Pterophyllum* species. The aligned ND4 sequence matrix of 649 bp consisted of 29 nucleotide sequences representative of the seven Neotropical outgroup taxa and 22 *Pterophyllum* taxa (**Addendum F**). A few gaps were introduced into the sequence matrix in order to accommodate a codon (CAA) unique to the *Pt. leopoldi* sequences. Furthermore, the "*Scalare* Genbank" sequence was significantly shortened at its 5' end (29.9% of its sequence length). However, as it was a reference sequence, it could not be re-sequenced and as such was not excluded from the ND4 sequence matrix. Similarly to the Cyt *b* gene, the ND4 gene also appeared to be a fast evolving gene. This was evident as the numerous variable characters across its short sequence length emphasized the nucleotide differences between the three *Pterophyllum* species and the outgroup taxa.

4.3.4.3 Nuclear matrices

The aligned RAG2 sequence matrix of 865 bp consisted of 30 nucleotide sequences representative of the seven Neotropical outgroup taxa and 23 *Pterophyllum* taxa (**Addendum G**). Aside from a few ambiguous base pairs within the nucleotide sequences of the outgroup taxa, no gaps had to be introduced in the sequence matrix in order to obtain optimal alignment of the 30 nucleotide sequences. The RAG2 gene appeared to be a slow evolving gene as there were many shared base pairs between the outgroup and *Pterophyllum* taxa. There were very few variable characters unique to each of the three *Pterophyllum* species. In contrast the aligned S7 sequence matrix of 618 bp consisted of 29 nucleotide sequences representative of the seven Neotropical outgroup taxa and 22 *Pterophyllum* taxa, consisting of only one *Pt. altum* representative from the Rio Inirida (**Addendum H**). Although not

significantly, the *Pt. leopoldi* D sequence was truncated at both its 3' and 5' ends so as to exclude ambiguous and nonsense base pairs as a consequence of poor sequencing. Furthermore, a few gaps (less than 5% of the aligned sequence length) were introduced in the sequence matrix in order to obtain optimal alignment thereof. The S7 region appeared to be a faster evolving region than the RAG2 gene as it contained more variable base pairs across its aligned sequence length.

The aligned Mitfb sequence matrix of 853 bp consisted of 19 nucleotide sequences representative of three Neotropical outgroup taxa (*Astronotus sp.*, *Heros severus* and *Symphysodon aequifasciatus*) and 16 *Pterophyllum* taxa excluding *Pt. scalare* species from Santa Isabel and Peru (**Addendum I**). As a consequence of including the three Neotropical outgroup taxa, a significant number of gaps (22% of aligned length) had to be introduced in the sequence matrix in order to obtain optimal alignment of the 19 taxa. The latter included compensating for a unique 18 bp nucleotide fragment shared only between the two outgroups *Heros severus* and *Symphysodon aequifasciatus*. Similarily, the aligned RpL8 sequence matrix of 861 bp consisted of 18 nucleotide sequences representative of the three same Neotropical outgroup taxa and 15 *Pterophyllum* taxa (**Addendum J**). The genus *Pterophyllum* excluded *Pt. scalare* representatives from Santa Isabel and Peru because insufficient DNA was supplied. In addition, the matrix only contained one representative from the Rio Inirida. However, only a few gaps (less than 5% of aligned sequence length) had to be introduced in order to achieve optimal alignment of the 18 nucleotide sequences. Mitfb appeared to be a faster evolving gene than RpL8 considering its aligned length and number of variable characters that are either shared or distinct between the nucleotide sequences representative of the three *Pterophyllum* species and in comparison to the three outgroup taxa.

4.3.5 Nucleotide Sequence Analysis And Alignment Of The Expanded ND4 Sequence Matrix

The expanded and aligned ND4 sequence matrix of 637 bp consisted of 51 nucleotide sequences representative of the seven Neotropical outgroup taxa and 44 *Pterophyllum* taxa. Optimal alignment of the expanded ND4 sequence matrix was obtained through the incorporation of several gaps. The latter involved the accommodation of a codon (CAA) unique to the *Pt. leopoldi* taxa. In addition, due to poor sequencing, the full length of the "Altum Rio Inirida 5" ND4 nucleotide sequence could not be recovered. Consequently, a nucleotide sequence fragment of 152 bp (positioned at base pairs 138 – 289) was missing which equates to 23.9% of the aligned sequence length. This was also true for the "Altum Rio Atabapo 4" sample. However, the latter was missing a nucleotide fragment of 91 bp (positioned at base pairs 294 – 384) which equates to 14.3% of the aligned sequence length. Unfortunately, due to limited sampling, the amplification of the ND4 gene from these two samples and the subsequent sequencing thereof could not be repeated. Nonetheless, the information contained within the expanded ND4 sequence matrix was regarded as sufficient to continue with the phylogenetic analysis.

4.3.6 PHYLOGENETIC ANALYSIS

4.3.6.1 Character variability assessment

A summary of the character variability, computed as a result of the heuristic search in a parsimony analysis, for each of the three reduced and aligned mitochondrial (Cyt *b*, 16S rRNA and ND4) sequence matrices is presented in **Table 4.4**.

TABLE 4.4 The character variability for each of the three reduced and aligned mitochondrial sequence matrices. Summarised are the total, constant, parsimony uninformative (PU) and parsimony informative (PI) characters for the three (Cyt *b*, 16S rRNA and ND4) reduced and aligned mitochondrial sequence matrices. In addition, the total number of variable (PU+PI) characters for each sequence matrix is calculated.

	MITOCHONDRIAL MARKERS					
	16S rRNA		Cyt *b*		ND4	
Total characters	562	-	1121	-	649	-
Constant	458	81.5%	705	62.9%	373	57.5%
Parsimony Uninformative (PU)	31	5.50%	109	9.70%	73	11.2%
Parsimony Informative (PI)	73	13.0%	307	27.4%	203	31.3%
Variable (PU + PI)	104	18.5%	416	37.1%	276	42.5%

A summary of the character variability, computed as a result of the heuristic search in a parsimony analysis, for each of the four aligned nuclear (RAG2, S7, Mitfb and RpL8) sequence matrices is presented in **Table 4.5**.

TABLE 4.5 The character variability for each of the four reduced and aligned nuclear sequence matrices. Summarised are the total characters, the number of constant, parsimony uninformative (PU), parsimony informative (PI) and the total number of variable (PU+PI) characters for each of the three aligned mitochondrial sequence matrices as computed by parsimony analysis.

	NUCLEAR MARKERS							
	RAG2		S7		Mitfb		RpL8	
Total characters	865	-	618	-	853	-	861	-
Constant characters	776	89.7%	491	79.4%	789	92.5%	750	87.1%
Parsimony Uninformative (PU)	44	5.10%	70	11.3%	39	4.60%	69	8.00%
Parsimony Informative (PI)	45	5.20%	57	9.20%	25	2.90%	42	4.90%
Variable characters (PU + PI)	89	10.3%	127	20.6%	64	7.50%	111	12.9%

A summary of the character variability for the concatenated sequence matrix consisting of the three mitochondrial (Cyt b, 16S rRNA and ND4) and two nuclear (RAG2 and S7) sequence matrices is presented in **Table 4.6**.

TABLE 4.6 The character variability for the reduced, concatenated and aligned five-gene sequence matrix. Summarised are the total characters, the number of constant, parsimony uninformative (PU), parsimony informative (PI) and the total number of variable (PU+PI) characters for each of the three aligned mitochondrial sequence matrices as computed by parsimony analysis.

	Combined character-status summaries				
	Mitochondrial (Cyt *b*, 16S rRNA, ND4)	Nuclear (RAG2, S7)	Total	% contribution	
				M	N
Total characters	2332	1483	3815	61.1	38.9
Constant characters	1536	1267	2803	54.8	45.2
Parsimony Uninformative (PU)	213	114	327	65.1	34.9
Parsimony Informative (PI)	583	102	685	85.1	14.9
Variable characters (PU + PI)	796	216	1012	78.7	21.3

The concatenated, reduced and aligned five-gene sequence matrix had a total length of 3,815 bp of which the three combined mitochondrial genes contributed 61.1% and the combined two nuclear genes 38.9%. Hence, based on total characters, the ratio of mitochondrial to nuclear data of the five-gene sequence matrix is roughly 3:2. Furthermore, parsimony analysis identified 2,803 unchanged base pairs among the total 35 nucleotide sequences analysed. Of the total constant characters, the three mitochondrial genes could account for 1,536 bp (54.8%) whilst the two nuclear markers represented the remaining 1,267 bp. Character variability assessed in the five-gene sequence matrix recovered 1,012 variable base pairs of which 78.7% was mitochondrial and 21.3% nuclear. Thus, in terms of the variable characters that define a data set, the ratio of mitochondrial to nuclear data is roughly 4:1. This means that the mitochondrial signal dominated the nuclear signal and as such determined the outcome of the five-gene analysis. A summary of the character variability for the reduced, concatenated and aligned seven-gene (Cyt *b*, 16S rRNA, ND4, RAG2, S7, Mitfb and RpL8) sequence matrix is presented in **Table 4.7**.

TABLE 4.7 The character variability for the reduced, concatenated and aligned seven-gene sequence matrix. Summarised are the total characters, the number of constant, parsimony uninformative (PU), parsimony informative (PI) and the total number of variable (PU+PI) characters for each of the three aligned mitochondrial sequence matrices as computed by parsimony analysis.

	Combined character-status summaries				
	Mitochondrial (Cyt *b*, 16S rRNA, ND4)	Nuclear (RAG2, S7, Mitfb, RpL8)	Total	% contribution	
				M	N
Total characters	2332	3197	5529	42.2	57.8
Constant characters	1536	2806	4342	35.4	64.6
Parsimony Uninformative (PU)	213	222	435	49.0	51.0
Parsimony Informative (PI)	583	169	752	77.5	22.5
Variable characters (PU + PI)	796	391	1187	67.1	32.9

In comparison with the concatenated, reduced and aligned five-gene sequence matrix, the seven-gene sequence matrix had a total length of 5,529 bp of which the difference of 1,714 bp can be attributed to the inclusion of the two nuclear markers Mitfb and RpL8. Of the total sequence length, the three combined mitochondrial genes contributed 42.2% and the combined four nuclear markers 57.8%. Hence, based on total characters, the ratio of mitochondrial to nuclear data of the seven-gene sequence matrix is roughly 2:3 which is exactly the inverse of what was observed for the five-gene sequence matrix. However, of the total constant characters, the three mitochondrial genes contributed 1,536 bp (35.4%) whilst the four nuclear markers contributed the remaining 2,806 bp which equates to 64.6%. Furthermore, character variability of the reduced seven-gene sequence matrix revealed a total of 1,187 variable base pairs which is equal to a substantial increase of 17.3% (175 bp) compared to the 1,012 variable sites of the reduced five-gene sequence matrix. Needless to say, this meaningful increase in variable characters can be ascribed to the additional two nuclear markers Mitfb and RpL8. Of the total 1,187 variable characters 67.1% was mitochondrial and 32.9% was nuclear. Thus, in terms of the variable characters, the ratio of mitochondrial to nuclear data is roughly 7:3. This means that even though the mitochondrial data represented roughly 40% of the sequence matrix, its signal yet again dominated the nuclear signal. As such, the mitochondrial data was the dominating contributor to the outcome of both the five- and seven-gene analyses. A summary of the tree scores for each of the seven molecular markers, the concatenated mitochondrial, concatenated nuclear and seven-gene sequence matrix are presented in **Table 4.8**.

TABLE 4.8 A summary of the tree scores for each of the 12 sequence matrices generated during the analysis. Summarised is the total number of trees, the length of the first tree of all the shortest possible trees and the CI and RI values for each of the 12 sequence matrices analysed.

	Cyt *b*	16S rRNA	ND4	Mitochondrial (3)	RAG2	S7
Total trees	1	1	290	9,780	1	2
CI	0.672	0.584	0.619	0.601	0.932	0.883
RI	0.821	0.776	0.803	0.786	0.956	0.916
Length #1	186	888	572	1,660	146	103
	Nuclear (2)	Mitfb	RpL8	Nuclear (4)	M+N (5)	M+N (7)
Total trees	30	2	1	7,900	9,770	8,570
CI	0.901	0.985	0.957	0.929	0.638	0.667
RI	0.932	0.983	0.945	0.940	0.799	0.806
Length #1	252	68	117	437	1,921	2,106

Likewise, the character variability of the expanded and aligned ND4 sequence matrix was computed of which the results are summarised in **Table 4.9**.

TABLE 4.9 The character variability for the expanded and aligned ND4 sequence matrix. Summarised are the total characters constituting the number of constant, parsimony uninformative (PU), parsimony informative (PI) and the total number of variable (PU+PI) characters as computed by parsimony analysis.

	ND4 expanded dataset	
Total characters	637	-
Constant characters	365	57.3%
Parsimony Uninformative (PU)	73	11.5%
Parsimony Informative (PI)	199	31.2%
Variable characters (PU + PI)	272	42.7%

Character variability of the expanded and aligned mitochondrial ND4 sequence matrix showed that 365 of the total 637 bp (57.3%) were constant whilst the remaining 272 bp (42.7%) were variable among the 51 taxa. Compared to the reduced and aligned ND4 sequence matrix, the expanded data set showed a minimal increase of 0.2% in the number of variable characters which can be ascribed to the slightly shorter sequence length (637 bp) and fewer variable characters (272 bp).

4.3.6.2 Maximum likelihood (ML) analysis

The three mitochondrial (Cyt *b*, 16S rRNA and ND4) gene trees are presented as **Figures 4.12** to **4.14** respectively. In spite of varying bootstrap support, the overall topology of the genus *Pterophyllum* remained the same throughout the three gene trees. However, one of the differences among the three mitochondrial gene trees was that each tree had a different topology for the *Pt. scalare* clade. The other difference among the three gene trees was the position of the genus *Pterophyllum* relative to its immediate neighbouring Heroini genera, more specifically the South American species *Hoplarchus psittacus, Hypselecara coryphaenoides* and the Mesonautines. The combined mitochondrial (Cyt *b*, 16S rRNA and ND4) gene tree is presented in **Figure 4.15**.

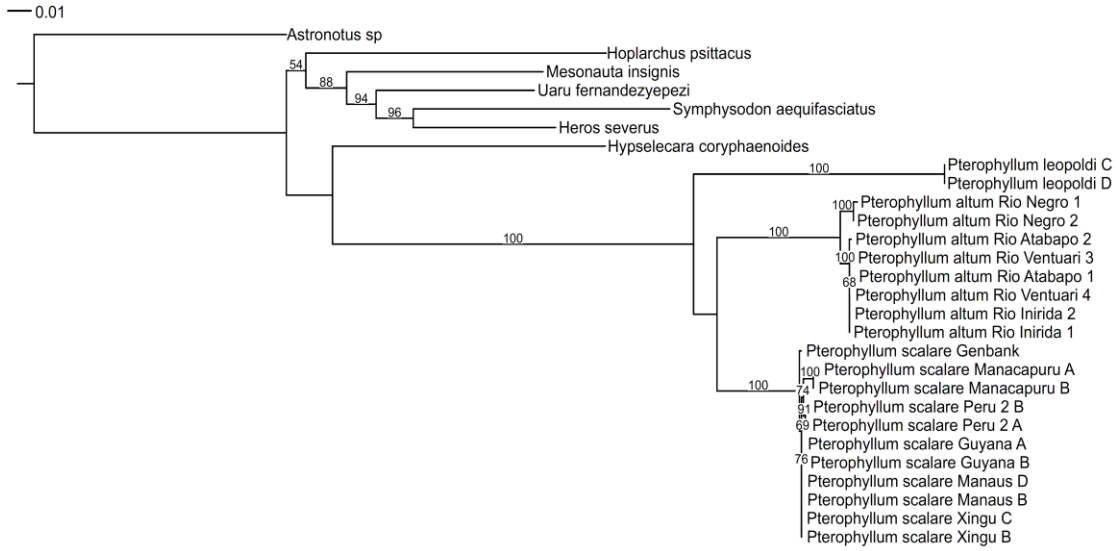


FIGURE 4.12 A bootstrap consensus tree of the reduced and aligned Cyt *b* sequence matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.

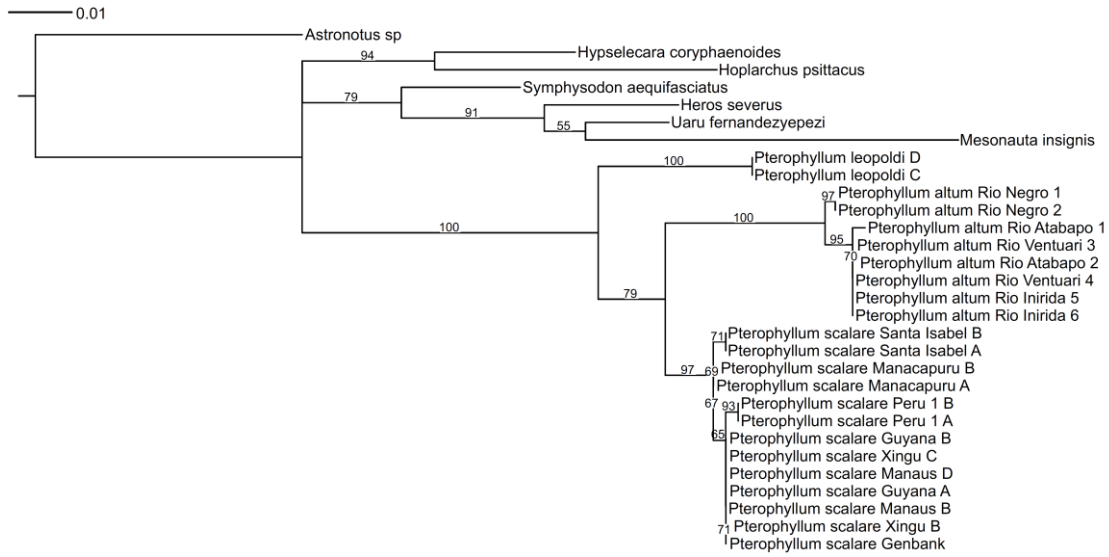


FIGURE 4.13 A bootstrap consensus tree of the reduced and aligned 16S rRNA sequence matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.

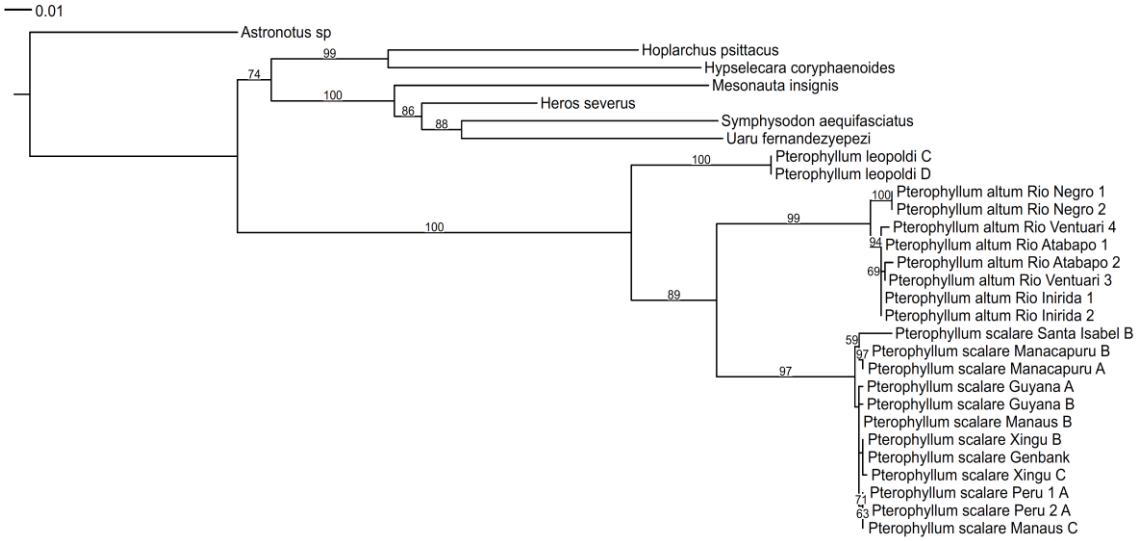


FIGURE 4.14 A bootstrap consensus tree of the reduced and aligned ND4 sequence matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.

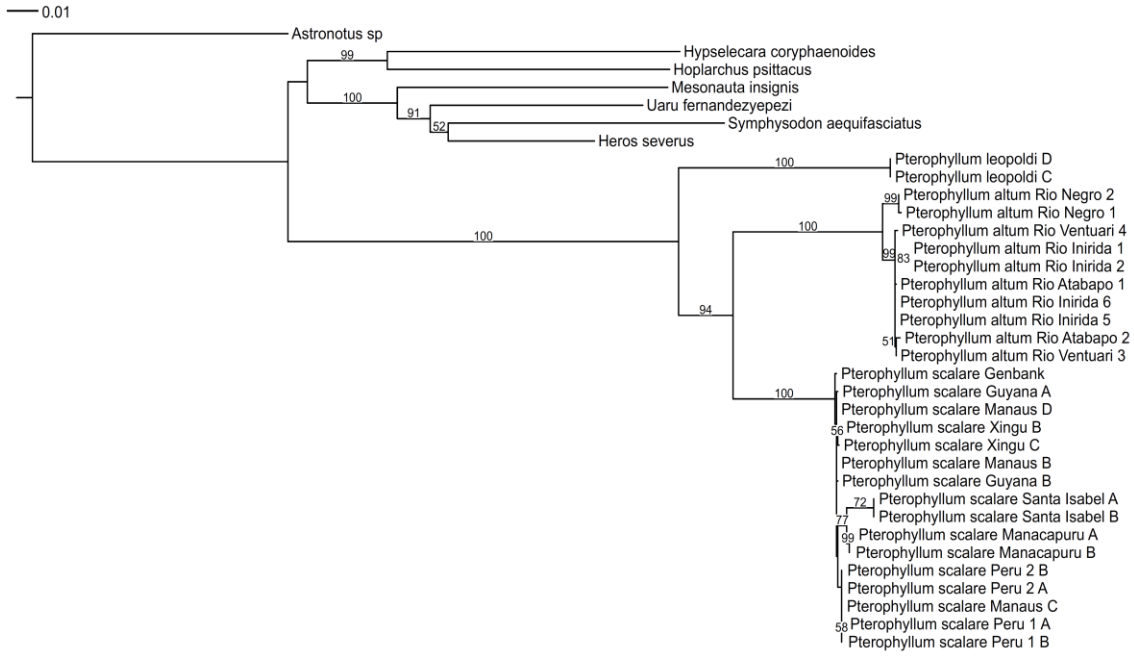


FIGURE 4.15 A bootstrap consensus tree of the reduced, concatenated and aligned mitochondrial (Cyt *b*, 16S rRNA, ND4) data matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.

The likelihood analysis performed on the concatenated mitochondrial sequence matrix retrieved the genus *Pterophyllum* as a monophyletic clade consisting of the three species *Pt. leopoldi*, *Pt. altum* and *Pt. scalare.* The *Pt. leopoldi* clade was recovered as the basal lineage and sister to the MRCA of the two closely related *Pt. altum* and *Pt. scalare* clades. Furthermore, a well-supported sister association was recovered between the *Pt. altum* representatives from the distinct upper Rio Negro and Rio Orinoco clades. Although not well resolved, three distinct groups (*Pt. scalare* Santa Isabel, *Pt. scalare* Manacapuru and *Pt. scalare* Peru) were recovered for *Pt. scalare*. In addition, as well-supported sister association was recovered between *Pt. scalare* representatives from Santa Isabel and Manacapuru. The genus *Pterophyllum* was retrieved in a sister association with the clade containing the two closely

related South American genera *Hoplarchus* and *Hypselecara* as well as the Mesonautines. Of the three mitochondrial genes, the ND4 gene tree was the most congruent with the combined mitochondrial results which is not suprising as it contributed a major proportion of the character variability.

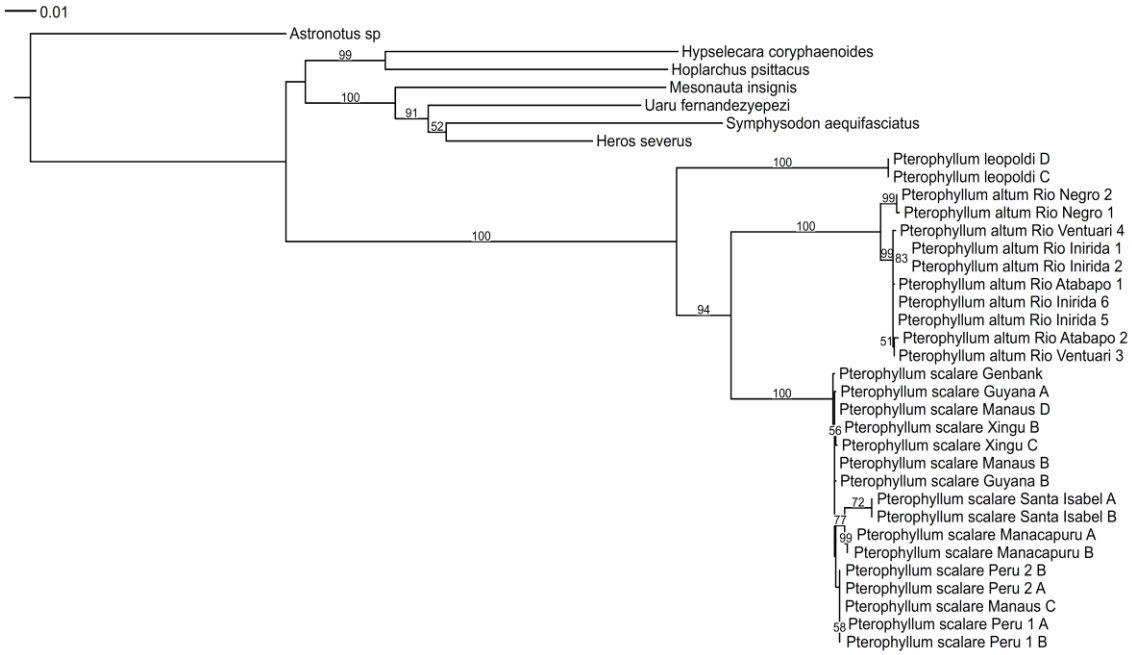


FIGURE 4.16 A bootstrap consensus tree of the reduced, concatenated and aligned mitochondrial (Cyt *b*, 16S rRNA, ND4) data matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.

The four nuclear (RAG2, S7, Mitfb and RpL8) gene trees are presented in **Figures 4.16** to **4.19** respectively. Although the four gene trees recovered the genus *Pterophyllum* to consist of the three species *Pt. leopoldi*, *Pt. altum* and *Pt. scalare*, the topology within the genus was different for each gene tree. The combined nuclear (RAG2, S7, Mitfb and RpL8) gene tree is presented in **Figure 4.20**.

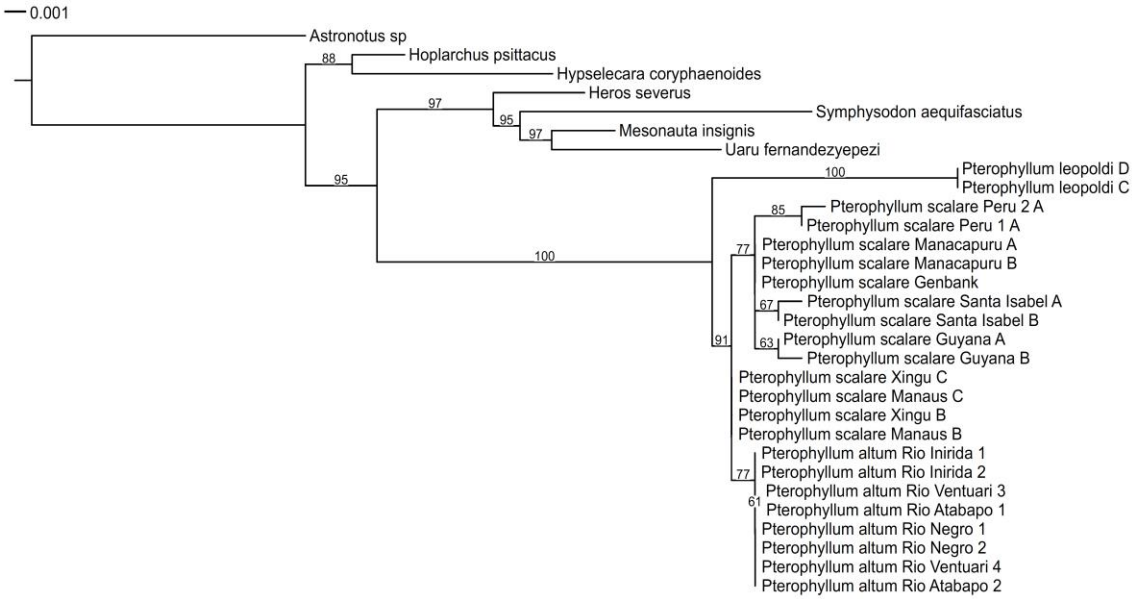


FIGURE 4.17 A bootstrap consensus tree of the reduced and aligned RAG2 sequence matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.

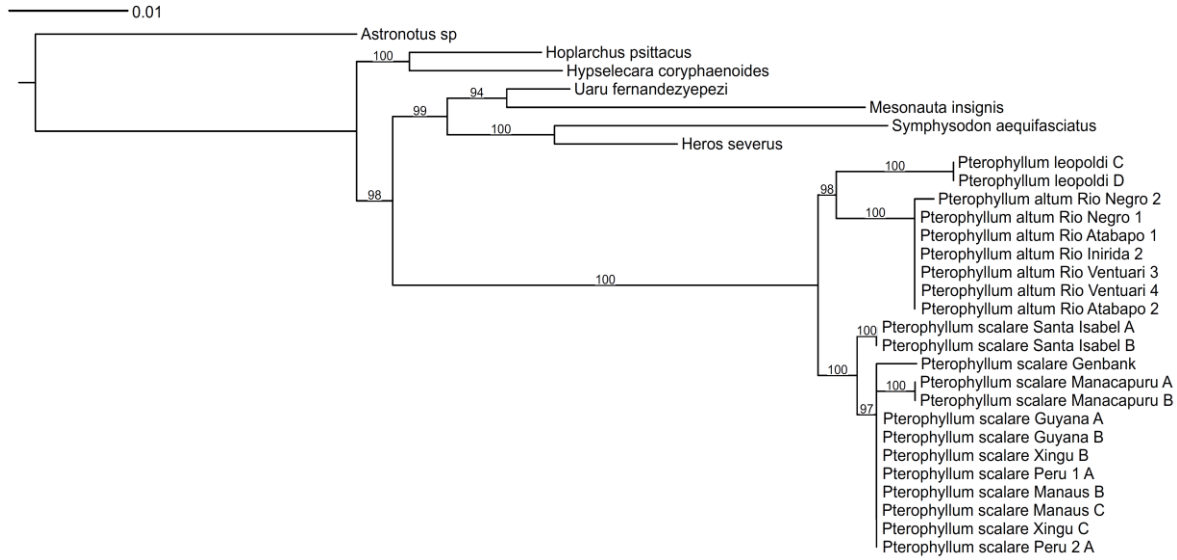


FIGURE 4.18 A bootstrap consensus tree of the reduced and aligned S7 sequence matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.

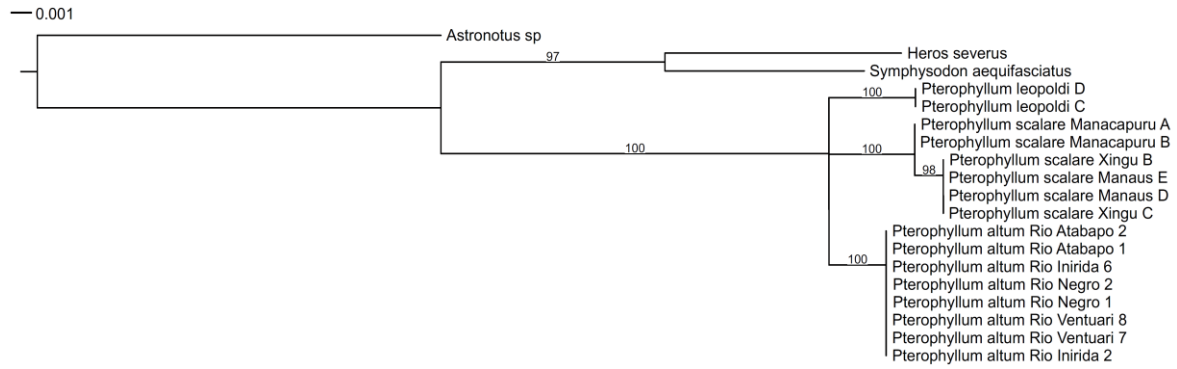


FIGURE 4.19 A bootstrap consensus tree of the reduced and aligned Mitfb sequence matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.

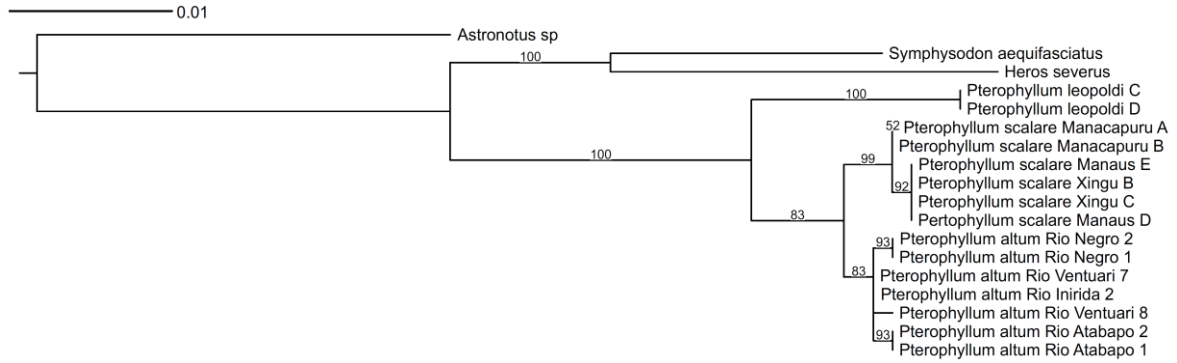


FIGURE 4.20 A bootstrap consensus tree of the reduced and aligned RpL8 sequence matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.

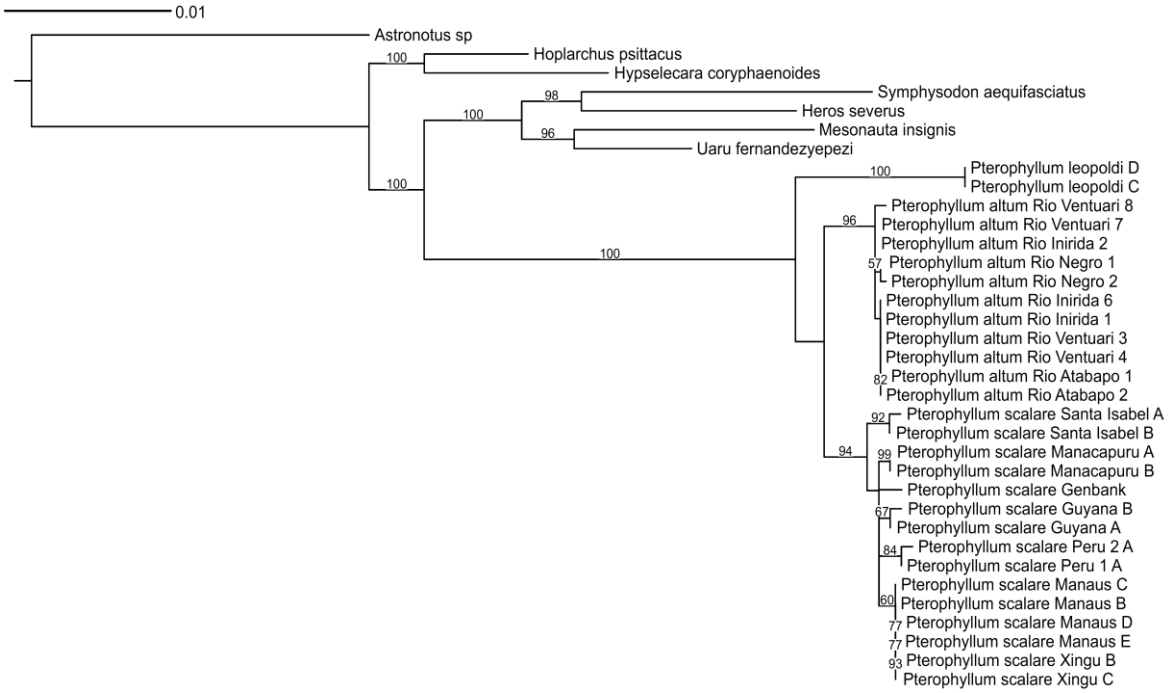


FIGURE 4.21 A bootstrap consensus tree of the reduced, concatenated and aligned nuclear (RAG2, S7, Mitfb and RpL8) sequence matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.

The likelihood analysis performed on the combined nuclear (RAG2, S7, Mitfb and RpL8) sequence matrix recovered the genus *Pterophyllum* as a monophyletic group. The *Pt. leopoldi* clade was recovered as the basal lineage and the *Pt. scalare* and *Pt. altum* clades recovered as sister groups. However, the *Pt. altum* clade was recovered as a polytomy without a clear distinction between the *Pt. altum* representatives from the upper Rio Negro and Rio Orinoco. Furthermore, the *Pt. scalare* clade was composed of two main groups; one containing the *Pt. scalare* representatives from Santa Isabel and the other, a polytomy, containing all other *Pt. scalare* representatives. The latter included the three distinct clades: *Pt. scalare* Manacapuru, *Pt. scalare* Guyana and *Pt. scalare* Peru. The genus *Pterophyllum* was retrieved as the sister clade of the monophyletic Mesonautines whilst the two closely related South American genera *Hoplarchus* and *Hypselecara* were recovered in a basal position.

The likelihood result of the combined seven-gene (Cyt *b*, 16S rRNA, ND4, RAG2, S7, Mitfb and RpL8) analysis is presented in **Figure 4.21**. The likelihood analysis performed on the concatenated seven-gene sequence matrix recovered the two South American genera *Hoplarchus* and *Hypselecara* as a monophyletic group and as the basal lineage of the Heroines. Furthermore, a sister association was recovered between the monophyletic Mesonautines and monophyletic genus *Pterophyllum*. Similarly to the concatenated mitochondrial and nuclear gene trees, the seven-gene analysis retrieved each of the three *Pterophyllum* species as monophyletic groups. In agreement with the mitochondrial data, the seven-gene analysis also recovered *Pt. leopoldi* as the basal lineage of the genus *Pterophyllum*. Moreover, the sister association recovered between *Pt. altum* and *Pt. scalare* was congruent with both the mitochondrial and nuclear data. The *Pt. altum* representatives from the upper

Rio Negro were retrieved as a well supported clade whilst the *Pt. altum* representatives from the Upper Rio Orinoco remained unresolved.

The *Pt. scalare* clade consisted of three distinct groups. The *Pt. scalare* Santa Isabel clade was retrieved as the basal lineage whilst the well supported *Pt. scalare* Manacapuru clade was sister to the remaining *Pt. scalare* representatives. Weakly supported *Pt. scalare* Peru and *Pt. scalare* Guyana clades were also retrieved.

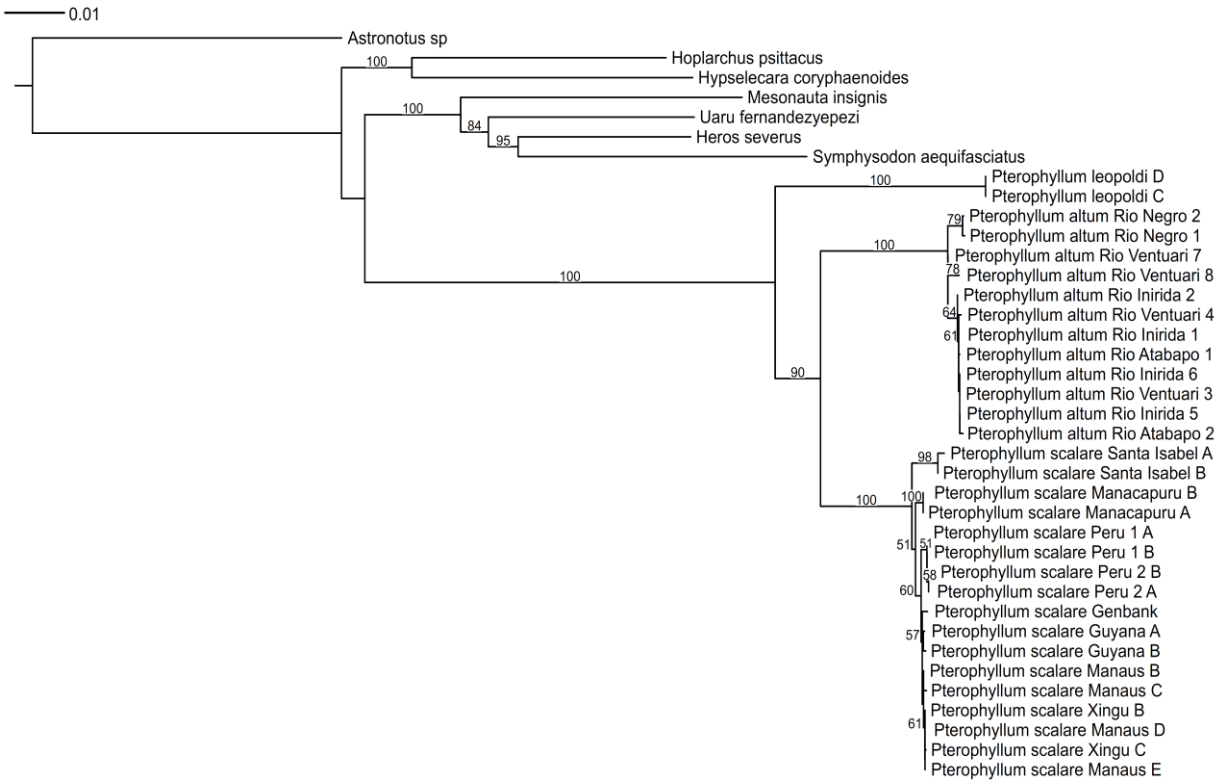


FIGURE 4.22 A bootstrap consensus tree of the reduced, aligned and concatenated seven-gene (Cyt *b*, 16S rRNA, ND4, RAG2, S7, Mitfb and RpL8) sequence matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.

The gene-tree retrieved from the likelihood analysis performed on the expanded ND4 sequence matrix is presented in **Figure 4.22**. Maximum likihood analysis recovered a similar topology as was obtained for the reduced and aligned ND4 sequence matrix. However, differences in the position of the taxa in the *Pt. scalare* and *Pt. altum* clades were evident. Within the *Pt. scalare* clade, two main groupings were recovered. One group included all Manacapuru taxa whilst the other group was recovered as a polytomy which included the remaining *Pt. scalare* samples. In addition, a sister association was recovered between the two taxa Santa Isabel B and *Scalare* Peru 1A. Although the *Pt. altum* clade was recovered as a polytomy, distinction can be made between the *Pt. altum* representatives from the upper Rio Negro and Rio Orinoco. No new insights into the variation present within *Pt. altum* representatives from the upper Rio Negro and Rio Orinoco could be drawn from the phylogeny based on the extended ND4 sequence matrix.

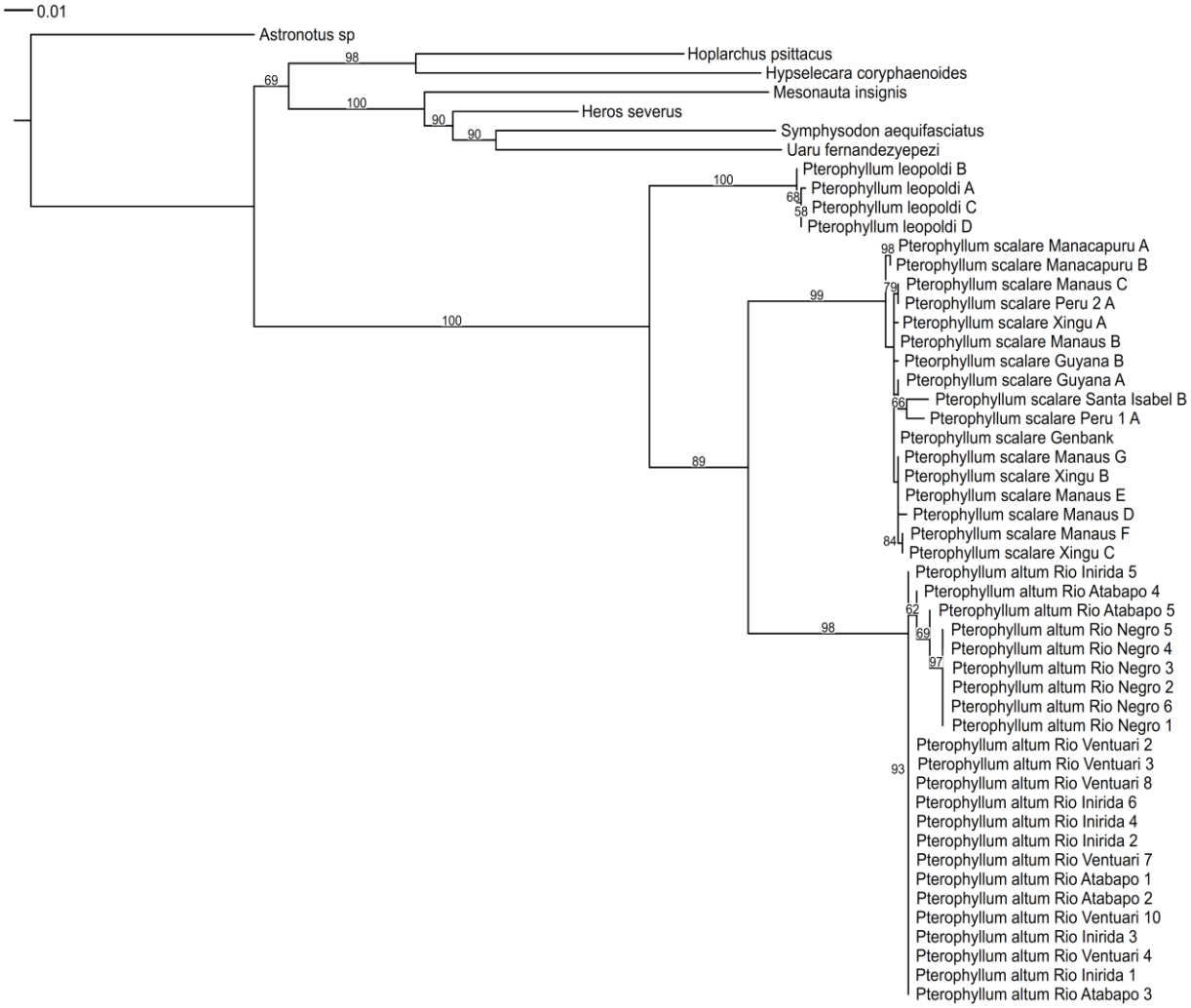


FIGURE 4.23 A bootstrap consensus tree of the reduced, aligned and extended ND4 sequence matrix. Branches are scaled proportionally and only bootstrap values equal and greater than 50% are annotated above branches.

4.4 DISCUSSION

In total 13 additional nuclear markers were investigated. The aim was to identify more variable nuclear markers to include in the phylogenetic analysis in order to increase the phylogenetic resolution within the genus *Pterophyllum*. Based on the amplification results and sequence analysis of the purified amplicons for each of the 13 nuclear markers, the latter were categorised into three groups. The first group of nuclear markers failed to amplify. This group included the two nuclear markers Sd15 and CteO12. The second group consisted of the three nuclear markers Sd11, Gpd2 and X-src of which the amplicons were not suitable for sequencing due to either a very low DNA concentration or multiple band formation. The third group consisted of the eight nuclear markers Endrb1, Sd23, GnRH3-3, Ib1, MpCS, TmoM27, Mitfb and RpL8. These nuclear markers were indeed successfully amplified and sequenced after which their sequence variability was investigated. The two most variable nuclear markers were found to be Mitfb and RpL8 and for this reason the sequences of these markers were determined for all samples so that they could be included in the phylogenetic analyses.

The reduced, concatenated and aligned five-gene sequence matrix showed that the three mitochondrial (Cyt *b*, 16S rRNA and ND4) versus the two nuclear (RAG2 and S7) markers strongly determined the outcome of the phylogenetic analysis. The total mitochondrial characters contributed 61.1% to the five-gene sequence matrix of which 78.7% were variable characters. The inclusion of the two additional nuclear markers Mitfb and RpL8 increased the total nuclear nucleotide characters from 38.9% to 57.8% in the reduced, concatenated and aligned seven-gene sequence matrix. However, the three mitochondrial markers still dominated (67.1%) the total variability of the seven-gene sequence matrix. The percentage of the variable nuclear characters increased from 21.3% to 32.9%. As such, the overall variability of the sequence matrix was increased.

Theoretically, this increased variability should result in improved resolution of the phylogeny within the genus *Pterophyllum.* However, if the resolution of the five-gene phylogeny, as presented in Chapter 3, is compared to the seven-gene phylogeny in Chapter 4, the resolution in the seven-gene phylogeny has weakened. This can be attributed to the reduced outgroup taxa from 166 Neotropical cichlids in the five-gene phylogeny to seven Heroini cichlids in the seven-gene phylogeny. This reduction of outgroups was made because the nuclear Mitfb and RpL8 genes of all the 166 Neotropical reference species could not be sequenced due to it being unavailable. According to Vandamme (2009) and Dowell (2008) increased taxon sampling of outgroups can potentially influence ingroup resolution and this is clearly evident here.

Resolution within both the *Pt. altum* and *Pt. scalare* clades in the seven-gene phylogeny has changed in comparison to the five-gene phylogeny. The *Pt. altum* clade no longer resolves into two distinct clades (Rio Negro and Rio Orinoco), but rather shows a basal polytomy. Hence, the reduction of outgroups weakened the resolution within the *Pt. altum* clade. However, resolution within the *Pt. scalare* clade improved. The *Pt. scalare* clade was retrieved as a basal polytomy in the five-gene phylogeny in which two emerging well-resolved clades (*Pt. scalare* Santa Isabel and *Pt. scalare* Manacapuru) were retrieved. However, in the seven-gene phylogeny, the positions of these two groups changed to a more basal position. Although resolution within the *Pt. scalare* clade improved, most likely due to the increased variability, the phylogenetic relationships were still not well-supported. However, the seven-gene phylogeny again confirmed that the genus *Pterophyllum* was a monophyletic group and confirmed that the genus consisted of three groups conforming to the presently accepted taxonomy of the three species i.e. *Pt. leopoldi*, *Pt. altum* and *Pt. scalare*. The species *Pt. leopoldi* was again retrieved as the basal lineage of the genus whilst *Pt. altum* and *Pt. scalare* formed a derived sister association. The expanded ND4 gene tree showed that very little intraspecies variation could be detected within the *Pt. altum* and *Pt. scalare* clades.

Based on the results of the seven-gene phylogeny the following conclusions can be drawn with regard to the taxa within the genus *Pterophyllum*. Firstly, there is no conflict between the mitochondrial and nuclear topologies. When taking into consideration the distance between the towns of Santa Isabel and

San Felipe (approximately 345 km by air and approximately 560 km by river), these results do not support hybridization between *Pt. altum* and *Pt. scalare* and therefore the hypothesis that *Pt. scalare* Santa Isabel has a hybrid origin must be rejected. Unfortunately, Meliciano could not investigate the possibility of hybridisation within taxa from the upper Rio Negro as her study was only based on mitochondrial Cyt *b* sequences and did not include nuclear data. However, based on her morphometric analyses, she noticed that fishes from Boa Vista and Santa Isabel were intermediates of *Pt. altum* and *Pt. scalare*. Within *Pt. scalare* the two resolved lineages *Pt. scalare* Santa Isabel and *Pt. scalare* Manacapuru were retrieved. This is consistent with the results obtained by Meliciano. Although her study was based solely on the mitochondrial Cyt *b* gene, significant differentiation was retrieved within *Pt. scalare*. This can be attributed to her extensive taxon sampling from various localities in the upper Rio Negro and Rio Amazonas.

5. CONCLUSION AND FUTURE PERSPECTIVES

Using the nucleotide sequences of five genes, three mitochondrial (Cyt *b*, 16S rRNA and ND4) and two nuclear (RAG2 and S7), in parsimony, maximum likelihood (ML) and Bayesian phylogenetic analyses, it was established that the genus *Pterophyllum* is monophyletic consisting of three species namely *Pterophyllum leopoldi*, *Pterophyllum altum* and *Pterophyllum scalare*. This study was the first to have included *Pt. altum* from its type locality i.e. Rio Atabapo. It was shown that the putative *Pt. altum* species from the upper Rio Negro are indeed of *Pt. altum* descent/origin and are distinct from *Pt. altum* of the Upper Rio Orinoco (Rio Atabapo, Rio Inirida and Rio Ventuari). Temporal divergence estimates retrieved for the three species of the genus *Pterophyllum* are congruent with the biogeographic split of the Amazon and Orinoco drainages as well as their recent reconnection via the Casiquiare Canal. As such, this study was the first to hypothesize a Pleistocene date (~0.79 Ma ago) for this reconnection between these two major South American drainages.

Inclusion of the two variable nuclear loci Mitfb and RpL8 in the taxonomically reduced and concatenated five-gene sequence matrix increased the number of informative nuclear characters, but due to reduced outgroup resolution, sequence variability was decreased. This did however lead to the recovery of the two weakly supported basal lineages *Pt. scalare* Santa Isabel and *Pt. scalare* Manacapuru within a monophyletic *Pt. scalare*. There was no evidence to support a hybrid origin of the Santa Isabel population with the putative parents *Pt. altum* and *Pt. scalare.* Increased taxon sampling in particular of *Pt. altum* and *Pt. scalare,* for expansion of the mitochondrial ND4 sequence matrix, could not uncover significant intraspecies variation within the three species groups.

In future, this investigation should be expanded by increased taxon sampling of *Pterophyllum* samples at population level from more extensive and representative collection localities in the tributaries of the Rio Amazonas, Rio Orinoco and Rio Essequibo. To this end, intraspecies relationships can be better elucidated and recently divergent intraspecies radiations/lineages identified. In accordance with this, the five genes (Cyt *b*, 16S rRNA, ND4, RAG2 and S7) as well as the two suitably variable nuclear markers Mitfb and RpL8 should be sequenced from the extensive *Pt. scalare* and *Pt. altum* samples as collected by Meliciano. Furthermore, the possible taxonomic implications of the phylogenetic positions of *Pt. scalare* Santa Isabel and *Pt. scalare* Manacapuru as basal lineages of *Pt. scalare* should be further investigated. The possibility of hybridisation occurring between *Pt. altum* and *Pt. scalare* species from the upper Rio Negro in populations such as the one of Santa Isabel should be further investigated using more suitable methods that give whole genome scans such as RAD Sequencing. In combination, these approaches should lead to a better understanding of the evolution of the genus *Pterophyllum* within the Neotropical Cichlidae.

REFERENCES

Albert, J.S., Reis, R.E., 2011. *Historical biogeography of Neotropical freshwater fishes*, 19th ed. University of California Press, California, United States.

Amado, M.V., Hrbek, T., Gravena, W., Fantin, C., DE Assunção, E.N., Astolfi-Filho, S., Farias, I.P., 2008. Isolation and characterization of microsatellite markers for the ornamental discus fish Symphysodon discus and cross-species amplification in other Heroini cichlid species. *Mol. Ecol. Resour*. **8**, 1451–3. doi:10.1111/j.1755-0998.2008.02200.x

Bleher, H., 2012. Aktuelles über "Blattfische": Arten und Formen der Skalare. *Amaz.* Süßwasseraquaristik - Fachmagazin **42**, 14–20.

Blenkinsopand, T., Moore, A., 2003. Tectonic geomorphology of passive margins continental interlands., in: Shroder, J., Owen, L.A. (Eds.), *Treatise on Geomorphology*. Academic Press, San Diego, California, USA, pp. 71–92.

Boumans, J., 2015. Finarama [WWW Document]. URL http://www.finarama.com/forum/viewtopic.php?t=1668 (accessed 1.1.15).

Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A.Y., Lim, Z.W., Bezault, E., Turner-Maier, J., Johnson, J., Alcazar, R., Noh, H.J., Russell, P., Aken, B., Alföldi, J., Amemiya, C., Azzouzi, N., Baroiller, J.-F., Barloy-Hubler, F., Berlin, A., Bloomquist, R., Carleton, K.L., Conte, M. a., D'Cotta, H., Eshel, O., Gaffney, L., Galibert, F., Gante, H.F., Gnerre, S., Greuter, L., Guyon, R., Haddad, N.S., Haerty, W., Harris, R.M., Hofmann, H. a., Hourlier, T., Hulata, G., Jaffe, D.B., Lara, M., Lee, A.P., MacCallum, I., Mwaiko, S., Nikaido, M., Nishihara, H., Ozouf-Costaz, C., Penman, D.J., Przybylski, D., Rakotomanga, M., Renn, S.C.P., Ribeiro, F.J., Ron, M., Salzburger, W., Sanchez-Pulido, L., Santos, M.E., Searle, S., Sharpe, T., Swofford, R., Tan, F.J., Williams, L., Young, S., Yin, S., Okada, N., Kocher, T.D., Miska, E. a., Lander, E.S., Venkatesh, B., Fernald, R.D., Meyer, A., Ponting, C.P., Streelman, J.T., Lindblad-Toh, K., Seehausen, O., Di Palma, F., 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature*. doi:10.1038/nature13726

Brown, T., 2007. *Genomes 3*, 3rd ed. Garland Science, New York.

Burress, E.D., 2015. Cichlid fishes as models of ecological diversification: patterns, mechanisms, and consequences. *Hydrobiologia* **748**, 7–27. doi:10.1007/s10750-014-1960-z

Chow, S., Hazama, K., 1998. Universal PCR primers for S7 ribosomal protein gene introns in fish. *Mol. Ecol*. **7**, 1255–6.

Concheiro Pérez, G.A., Rícan, O., Ortí, G., Bermingham, E., Doadrio, I., Zardoya, R., 2007. Phylogeny and biogeography of 91 species of heroine cichlids (Teleostei: Cichlidae) based on sequences of the cytochrome *b* gene. *Mol. Phylogenet. Evol.* **43**, 91–110. doi:10.1016/j.ympev.2006.08.012

Darriba, D., Taboada, G., Doallo, R., Posada, D., 2012. jModelTest2: more models, new heuristics and parallel computing. *Nat. Methods* **9,** 772.

Davey, J.W., Blaxter, M.L., 2010. RADSeq: next-generation population genetics. *Brief. Funct. Genomics* **9**, 416–423. doi:10.1093/bfgp/elq031

Dorn, A., Ng'oma, E., Janko, K., Reichwald, K., Polačik, M., Platzer, M., Cellerino, A., Reichard, M., 2011. Phylogeny, genetic variability and colour polymorphism of an emerging animal model: the short-lived annual Nothobranchius fishes from southern Mozambique. *Mol. Phylogenet. Evol.* **61**, 739–49. doi:10.1016/j.ympev.2011.06.010

Dowell, K., 2008. Molecular Phylogenetics:An introduction to computational methods and tools for analyzing evolutionary relationships. Orono, Maine.

Drummond, A., Ho, S., Rawlence, N., Rambaut, A., 2007. A Rough Guide to BEAST 1 . 4.

Drummond, A., Rambaut, A., 2009. Bayesion evolusionary analysis by sampling trees, in: Lemmey, P., Salemi, M., Vandamme, A. (Eds.), *The Phylogenetic Handbook*. Cambridge University Press, Cambridge, UK, pp. 564–575.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC *Evol. Biol*. **7**, 214. doi:10.1186/1471-2148-7-214

Drummond, A.J., Suchard, M. A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29,** 1969–73. doi:10.1093/molbev/mss075

Farias, I.P., Ortí, G., Meyer, A., 2000. Total evidence: molecules, morphology, and the phylogenetics of cichlid fishes. *J. Exp. Zool*. **288**, 76–92.

Farias, I.P., Ortí, G., Sampaio, I., Schneider, H., Meyer, A., 2001. The cytochrome *b* gene as a phylogenetic marker: the limits of resolution for analyzing relationships among cichlid fishes. *J. Mol. Evol*. **53**, 89–103. doi:10.1007/s002390010197

Farias, I.P., Ortí, G., Sampaio, I., Schneider, H., Meyer, A., 1999. Mitochondrial DNA phylogeny of the family Cichlidae: monophyly and fast molecular evolution of the neotropical assemblage. *J. Mol. Evol*. **48**, 703–11.

Felsenstein, J., 1985. Confidence limits on phylogenies: An approach using Bootstrap. *Evolution* (N. Y). **39**, 783–791.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: A maximum-likelihood approach. *J. Mol. Evol.* **17**, 368–376.

Fisher-Reid, M.C., Wiens, J.J., 2011. What are the consequences of combining nuclear and mitochondrial data for phylogenetic analysis? Lessons from Plethodon salamanders and 13 other vertebrate clades. *BMC Evol. Biol.* **11**, 300. doi:10.1186/1471-2148-11-300

Forkel, S., 2015. Scalarezucht.de [WWW Document]. URL http://www.skalarezucht.de/ (accessed 10.8.15).

Friedman, M., Keck, B.P., Dornburg, A., Eytan, R.I., Martin, C.H., Hulsey, C.D., Wainwright, P.C., Near, T.J., 2013. Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. *Proc. R. Soc. B Biol. Sci.* **280**, 20131733. doi:10.1098/rspb.2013.1733

Genner, M.J., Seehausen, O., Lunt, D.H., Joice, D.A., Shaw, P.W., Carvalho, G., Turner, G., 2007. Age of cichlids: new dates for ancient lake fish radiations. *Mol. Biol. Evol.* **24**, 1269–1282.

Gernhard, T., 2008. The conditioned reconstructed process. *J. Theor. Biol.* **253**, 769–78. doi:10.1016/j.jtbi.2008.04.005

Guidon, S., Gascuel, O., 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Syst. Biol.* **52**, 696–704.

Hall, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98.

Hao, G., Wu, Q., Zhong, H., Zhou, Y., 2015. Complete mitochondrial genome of Pterophyllum scalare (Perciformes, Cichlidae). *Mitochondrial DNA* 1–2. doi:10.3109/19401736.2015.1022746

Hartl, D., Clark, A., 2007. Molecular population genetics, in: *Principles of Population Genetics*. Sinauer Associates, Sunderland, Massachusetts, USA, pp. 362–363.

Hassan, M., Lemaire, C., Fauvelot, C., Bonhomme, F., 2002. Seventeen new exon-primed intron-crossing polymerase chain reaction amplifiable introns in fish. *Mol. Ecol. Notes* **2**, 334–340. doi:10.1046/j.1471-8286

Hoorn, C., Guerrero, J., Sarmiento, G. A., Lorente, M. A., 1995. Andean tectonics as a cause for changing drainage patterns in Miocene northern South America. *Geology* **23**, 237–240. doi:10.1130/0091-7613(1995)023<0237:ATAACF>2.3.CO;2

Hoorn, C., Wesselingh, F.P., 2010. *Amazonia, Landscape and Species Evolution: A Look into the Past*, 1st ed. Wiley-Blackwell Publishers, Chichester, West Sussex, PO19 8SQ, UK.

Hoorn, C., Wesselingh, F.P., Steege, H., Bermudez, M. A., Mora, A., Sevink, J., Sanmartín, I., Anderson, C.L., Figueiredo, J.P., Jaramillo, C., Riff, D., 2010. Amazonia Through Time : Andean. *Science* **330**, 927–931. doi:10.1126/science.1194585

Hulsey, C.D., Keck, B.P., Hollingsworth, P.R., 2011. Species tree estimation and the historical biogeography of heroine cichlids. *Mol. Phylogenet. Evol*. **58**, 124–31. doi:10.1016/j.ympev.2010.11.016

Hulsey, D., García de León, F.J., Sánchez Johnson, Y., Hendrickson, D. A., Near, T.J., 2004. Temporal diversification of Mesoamerican cichlid fishes across a major biogeographic boundary. *Mol. Phylogenet. Evol*. **31**, 754–64. doi:10.1016/j.ympev.2003.08.024

Ilves, K.L., López-Fernández, H., 2014. A targeted next-generation sequencing toolkit for exon-based cichlid phylogenomics. *Mol. Ecol. Resour*. **14,** 802–811. doi:10.1111/1755-0998.12222

Joyce, D.A., Lunt, D.H., Genner, M.J., Turner, G.F., Bills, R., Seehausen, O., 2011. Erratum: Repeated colonization and hybridization in Lake Malawi cichlids (Current Biology (2011) 21 (R108-R109)). *Curr. Biol.* **21**, 526. doi:10.1016/j.cub.2011.02.044

Karl, S.A., Streelman, J.T., 1997. Reconstructing Labroid Evolution with Single-Copy Nuclear DNA. *Proc. Biol. Sci*. **264**, 1011–1020.

Keller, I., Wagner, C.E., Greuter, L., Mwaiko, S., Selz, O.M., Sivasundar, A., Wittwer, S., Seehausen, O., 2013. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Mol. Ecol*. **22**, 2848–2863. doi:10.1111/mec.12083

Klug, W.S., Cummings, M.R., Spencer, C.A., Palladino, M.A., 2009. *Concepts of Genetics*, 9th ed. Pearson Benjamin Cummings, San Fransico, USA.

Kocher, T.D., Stepien, C.A. (Eds.), 1997. *Molecular Systematics of Fishes*. Academic Press, San Diego, California, USA.

Kullander, S.O., 2003. Family Cichlidae (Cichlids), in: *Check List of the Freshwater Fishes of South and Central America*. Edipucrs, Porto Alegre, Brazil, pp. 605–654.

Kullander, S.O., 1983. Taxonomic studies on the Percoid freshwater fish family Cichlidae in South America. Stockholms Universitet, Stockholm, Sweden.

Kullander, S.O., 1998. A phylogeny and classification of the South American Cichlidae (Teleostei: Perciformes), in: Malabarba, L.R., Reis, R.E., Vari, R.P., Lucena, Z.M., Lucena, C.A.S. (Eds.), *Phylogeny and Classification of Neotropical Fishes*. Part 5 - Perciformes. Edipucrs, Porto Alegre, pp. 461–498.

Kullander, S.O., 1986. *Cichlid fishes of the Amazon River drainage of Peru*. Swedish Museum of Natural History, Stockholm, Sweden.

Letunic, I., Bork, P., 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res*. **39**, W475–W478. doi:10.1093/nar/gkr201

Letunic, I., Bork, P., 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128. doi:10.1093/bioinformatics/btl529

Lévêque, C., Oberdorff, T., Paugy, D., Stiassny, M.L.J., Tedesco, P. A., 2007. Global diversity of fish (Pisces) in freshwater. *Hydrobiologia* **595**, 545–567. doi:10.1007/s10750-007-9034-0

Li, C., Riethoven, J.-J.M., Ma, L., 2010. Exon-primed intron-crossing (EPIC) markers for non-model teleost fishes. *BMC Evol. Biol.* **10**, 90. doi:10.1186/1471-2148-10-90

López-Fernández, H., Arbour, J.H., Winemiller, K.O., Honeycutt, R.L., 2013. Testing for ancient adaptive radiations in neotropical cichlid fishes. *Evolution* **67**, 1321–37. doi:10.1111/evo.12038

Lopez-Fernandez, H., Honeycutt, R.L., Stiassny, M.L.J., Winemiller, K.O., 2005. Morphology , molecules , and character congruence in the phylogeny of South American geophagine cichlids (Perciformes , Labroidei). *Zool. Scr*. **34**, 627–651. doi:10.1111/j.1463-6409.2005.00209.x

López-Fernández, H., Honeycutt, R.L., Winemiller, K.O., 2005. Molecular phylogeny and evidence for an adaptive radiation of geophagine cichlids from South America (Perciformes: Labroidei). *Mol. Phylogenet. Evol*. **34**, 227–44. doi:10.1016/j.ympev.2004.09.004

López-Fernández, H., Winemiller, K.O., Honeycutt, R.L., 2010. Multilocus phylogeny and rapid radiations in Neotropical cichlid fishes (Perciformes: Cichlidae: Cichlinae). *Mol. Phylogenet. Evol*. **55**, 1070–86. doi:10.1016/j.ympev.2010.02.020

Margush, T., 1981. Consensus n-trees. *Bull. Math. Biol*. **43**, 239–244.

McMahan, C.D., Chakrabarty, P., Sparks, J.S., Smith, W.M.L., Davis, M.P., 2013. Temporal patterns of diversification across global cichlid biodiversity (Acanthomorpha: Cichlidae). *PLoS One* **8**, e71162. doi:10.1371/journal.pone.0071162

Miller, M. A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 *Gatew. Comput. Environ. Work.* 1–8. doi:10.1109/GCE.2010.5676129

Murray, A.M., 2001. The fossil record and biogeography of the Cichlidae (Actinopterygii: Labroidei). *Biol. J. Linn. Soc*. **74**, 517–532. doi:10.1006/bijl.2001.0599

Murray, A.M., 2000. Eocene Cichlid Fishes from Tanzania , East Africa. *J. Vertebr. Paleontol*. **20**, 651–664.

Musilová, Z., Rícan, O., Janko, K., Novák, J., 2008. Molecular phylogeny and biogeography of the Neotropical cichlid fish tribe Cichlasomatini (Teleostei: Cichlidae: Cichlasomatinae). *Mol. Phylogenet. Evol*. **46**, 659–72. doi:10.1016/j.ympev.2007.10.011

Palumbi, S., Martin, A., Romano, S., McMillan, W.O., Stice, L., Grabowski, G., 1991. The Simple Fool's Guide to PCR. Univ. Hawaii, Honolulu.

Pathak, D., Ali, S., 2012. Repetitive DNA : A Tool to Explore Animal Genomes / Transcriptomes, in: Meroni, G., Petrera, F. (Eds.), Functional Genomics. *InTech.* doi:10.5772/48259

Pereira, S.L., 2000. Mitochondrial genome organization and vertebrate phylogenetics. *Genet. Mol. Biol*. **23**, 745–752. doi:10.1590/S1415-47572000000400008

Platt, A., Woodhall, R., George Jr, A., 2007. Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol. *Biotechniques* **43**, 58–62. doi:10.2144/000112499

Rambaut, A., Suchard, M., Xie, D., Drummond, A., 2014. Tracer v1.6 [WWW Document]. URL http://beast.bio.ed.ac.uk/Tracer (accessed 10.13.14).

Říčan, O., Piálek, L., Zardoya, R., Doadrio, I., Zrzavý, J., 2013. Biogeography of the Mesoamerican Cichlidae (Teleostei: Heroini): Colonization through the GAARlandia land bridge and early diversification. *J. Biogeogr*. **40**, 579–593. doi:10.1111/jbi.12023

Rícan, O., Zardoya, R., Doadrio, I., 2008. Phylogenetic relationships of Middle American cichlids (Cichlidae, Heroini) based on combined evidence from nuclear genes, mtDNA, and morphology. *Mol. Phylogenet. Evol*. **49**, 941–57. doi:10.1016/j.ympev.2008.07.022

Roe, K.J., Conkel, D., Lydeard, C., 1997. Molecular systematics of Middle American cichlid fishes and the evolution of trophic-types in "Cichlasoma (Amphilophus)" and "C. (Thorichthys)". *Mol. Phylogenet. Evol.* **7**, 366–376. doi:10.1006/mpev.1997.0408

Rogers, J.J., Santosh, M., 2004. Gondwana and Pangea, in: Continents and Supercontinents. p. 114.

Schmidt, Heiko, A., von Haeseler, A., 2009. Phylogenetic inference using maximum likelihood methods, in: Lemey, P., Salemi, M., Vandamme, A.-M. (Eds.), *The Phylogenetic Handbook*. Cambridge University Press, Cambridge, UK, pp. 181–185.

Schneider, C.H., Gross, M.C., Terencio, M.L., Artoni, R.F., Vicari, M.R., Martins, C., Feldberg, E., 2012. Chromosomal evolution of neotropical cichlids: the role of repetitive DNA sequences in the organization and structure of karyotype. *Rev. Fish Biol. Fish.* **23,** 201–214. doi:10.1007/s11160-012-9285-3

Sides, J., Lydeard, C., 2000. Phylogenetic utility of the tyrosine kinase gene X-src for assessing relationships among representative cichlid fishes. *Mol. Phylogenet. Evol.* **14**, 51–74. doi:10.1006/mpev.1999.0693

Simons, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* doi:10.1080/17470210902990829

Smith, W., Chakrabarty, P., Sparks, J., 2008. Phylogeny, taxonomy and evolution of Neotropical cichlids (Teleostei: Cichlidae: Cichlinae). *Cladistics* **24**, 625–641.

Sparks, J., Smith, W., 2004. Phylogeny and biogeography of cichlid fishes (Teleostei: Perciformes: Cichlidae). *Cladistics* **20**, 501–517.

Sparks, J.S., 2004. Molecular phylogeny and biogeography of the Malagasy and South Asian cichlids (Teleostei: Perciformes: Cichlidae). *Mol. Phylogenet. Evol*. **30**, 599–614. doi:10.1016/S1055-7903(03)00225-2

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–3. doi:10.1093/bioinformatics/btu033

Steele, S.E., López-Fernández, H., 2014. Body size diversity and frequency distributions of Neotropical cichlid fishes (Perciformes: Cichlidae: Cichlinae). *PLoS One* **9**, 1–11. doi:10.1371/journal.pone.0106336

Stiassny, M.L.J., 1991. Phylogenetic intrarelationships of the family Cichlidae : an overview, in: Keenleyside, M.H.A. (Ed.), *Cichlid Fishes - Behaviour, Ecology and Evolution*. Chapman and Hall, London, pp. 1–33.

Stiassny, M.L.J., 1987. Cichlid familial intrarelationships and the placement of the neotropical genus Cichla (Perciformes, Labroidei). *J. Nat. Hist*. **21**, 1311–1331. doi:10.1080/00222938700770811

Streelman, J.T., Zardoya, R., Meyer, A., Karl, S. A., 1998. Multilocus phylogeny of cichlid fishes (Pisces: Perciformes): evolutionary comparison of microsatellite and single-copy nuclear loci. *Mol. Biol. Evol*. **15**, 798–808.

Suchard, M. A., Rambaut, A., 2009. Many-core algorithms for statistical phylogenetics. *Bioinformatics* **25**, 1370–6. doi:10.1093/bioinformatics/btp244

Swofford, D., 2003. PAUP: Phylogenetic analysis using parsimony (*and other methods). Vers. 4.0b10.

Swofford, D., Sullivan, J., 2009. Phylogeny inference based on parsimony and other methods using PAUP*, in: Lemey, P., Salemi, M., Vandamme, A.-M. (Eds.), *The Phylogenetic Handbook.* Cambridge University Press, Cambridge, UK, pp. 267–270.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. **22**, 4673–80.

Vaidya, G., Lohman, D.J., Meier, R., 2011. Cladistics multi-gene datasets with character set and codon information **27**, 171–180.

Vandamme, A.M., 2009. Basic Concepts of Molecular Evolution, in: Lemey, P., Salemi, M., Vandamme, A.-M. (Eds.), *The Phylogenetic Handbook.* Cambridge University Press, Cambridge, UK, pp. 19–23.

Willis, S.C., 2011. Multilocus and parametric analyses of the evolutionary history of the Amazonian peacock cichlids , the genus Cichla (Teleostei : Cichlidae). University of Nebraska.

Willis, S.C., 2015. Evolution and Ecology of Fishes [WWW Document]. URL https://sites.google.com/site/stuartcwillis/uaupes (accessed 12.17.15).

Willis, S.C., Nunes, M., Montaña, C.G., Farias, I.P., Ortí, G., Lovejoy, N.R., 2010. The Casiquiare river acts as a corridor between the Amazonas and Orinoco river basins: biogeographic analysis of the genus Cichla. *Mol. Ecol.* **19**, 1014–30. doi:10.1111/j.1365-294X.2010.04540.x

Willis, S.C., Nunes, M.S., Montaña, C.G., Farias, I.P., Lovejoy, N.R., 2007. Systematics, biogeography, and evolution of the Neotropical peacock basses Cichla (Perciformes: Cichlidae). *Mol. Phylogenet. Evol.* **44**, 291–307. doi:10.1016/j.ympev.2006.12.014

Winemiller, K.O., López-Fernández, H., Taphorn, D.C., Nico, L.G., Duque, A.B., 2008. Fish assemblages of the Casiquiare River, a corridor and zoogeographical filter for dispersal between the Orinoco and Amazon basins. *J. Biogeogr.* **35**, 1551–1563. doi:10.1111/j.1365-2699.2008.01917.x

Winemiller, K.O., Willis, S.C., 2011. The Vaupes Arch and Casiquiare Canal; Barriers and Passages, in: Albert, J.S., Reis, R.E. (Eds.), *Historical Biogeography of Neotropical Freshwater Fishes.* University of California Press, pp. 225–242.

Won, Y.J., Wang, Y., Sivasundar, A., Raincrow, J., Hey, J., 2006. Nuclear gene variation and molecular dating of the cichlid species flock of Lake Malawi. *Mol. Biol. Evol*. **23**, 828–37. doi:10.1093/molbev/msj101

Zardoya, R., Meyer, A., 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Mol. Biol. Evol.* **13**, 933–942. doi:10.1093/oxfordjournals.molbev.a025661

Zardoya, R., Vollmer, D.M., Craddock, C., Streelman, J.T., Karl, S., Meyer, A., 1996a. Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). *Proc. Biol. Sci.* **263**, 1589–1598. doi:10.1098/rspb.1996.0233

Zardoya, R., Vollmer, D.M., Craddock, C., Streelman, J.T., Karl, S., Meyer, A., 1996b. Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). *Proc. Biol. Sci.* **263**, 1589–98. doi:10.1098/rspb.1996.0233

ADDENDUM A

A detailed desription of the buffers used in this study. Summarised are the constituents of each of the buffers used during electrophoresis and PCR product purification.

Buffer	Constituents
1 X Casting buffer	20 mL 50X TAE electrophoresis buffer 980 mL Milli-Q®
1 X Running buffer	20 mL 50X TAE electrophoresis buffer 980 mL RO water
50X TAE electrophoresis buffer	0.48% (w/v) Tris-base 0.11% (v/v) glacial acetic acid 0.5 M EDTA at pH 8.0
Loading buffer	0.25% (w/v) bromophenol blue 57.5% (v/v) glycerol 0.5 M EDTA 1 M Tris pH 8.0 water

ADDENDUM B

A summary of all the outgroup and Neotropical reference sequences used to generate a five-gene phylogeny. The table summarizes the GenBank accession numbers for each of the 166 nucleotide sequences that apply to three mitochondrial- and two nuclear markers. These nucleotide sequences are representative of six outgroup and 160 Neotropical reference taxa (López-Fernández et al., 2010).

	GENBANK ACCESSION NUMBERS				
	Mitochondrial			**Nuclear**	
Taxon	Cyt *b*	ND4	16S rRNA	RAG2	S7
Outgroups					
Chromidotilapia guntheri	GU736923	GU736995	GU737111	-	GU736677
Etroplus maculatus	AF370625	-	GU737110	AY279874	DQ119279
Hemichromis fasciatus	AY050618	GU736996	GU737112	-	GU736678
Heterochromis multidens	AF370636	GU736993	GU737109	GU736811	GU736676
Paratilapia polleni	AF370627	GU736994	DQ119193	-	DQ119280
Paretroplus polyactis	GU736924	-	GU737113	-	GU736679
Astronotini					
Astronotus sp.	GU736926	AY566776	GU737117	AY566740	GU736683
Chaetobranchini					
Chaetobranchus flavescens	AF370652	GU736998	GU737118	GU736812	GU736684
Chaetobranchopsis orbicularis	AF370651	GU736999	AY662728	-	-
Cichlasomatini					
Acaronia nassa	GU736974	GU737029	GU737184	GU736845	GU736750

	GENBANK ACCESSION NUMBERS				
	Mitochondrial			Nuclear	
Taxon	Cyt *b*	ND4	16S rRNA	RAG2	S7
Acaronia vultuosa	GU736975	GU737030	GU736975	GU736846	GU736751
Aequidens diadema	AY050611	GU737015	GU737168	GU736828	GU736732
Aequidens tetramerus	AY050609	GU737016	GU737169	GU736829	GU736733
Aequidens tetramerus 'Xingu'	GU736963	GU737017	GU737170	GU736830	GU736734
Andinoacara coeruleopunctatus	AY843377	GU737019	GU737172	GU736832	GU736736
Andinoacara pulcher	AY050617	GU737021	AY294128	GU736834	GU736738
Andinoacara rivulatus	GU736965	GU737020	GU737173	GU736833	GU736737
Bujurquina aff. apoparuana	GU736966	-	GU737174	GU736835	GU736739
Bujurquina aff. hophrys	GU736967	-	GU737175	GU736836	GU736740
Cichlasoma amazonarum	AF370669	GU737023	AF045845	GU736838	GU736743
Cichlasoma dimerus	GU736969	GU737022	GU737177	GU736837	GU736742
Cichlasoma orinocense	GU736968	AY566778	GU737176	AY566747	GU736741
Cleithracara maronii	GU050614	GU737024	GU737178	GU736839	GU736744
Krobia potaroensis	GU736964	GU737018	GU737171	GU736831	GU736735
Krobia sp. 'Xingu orange spot'	GU736970	GU737025	GU737179	GU736840	GU736745
Laetacara thayeri	AY050608	GU737026	GU737180	GU736841	GU736746
Laetacara dorsigera	GU736971	GU737027	GU737181	GU736842	GU736747
Nannacara taenia	GU736972	-	GU737182	GU736843	GU736748
Tahuantinsuyoa macantzatza	GU736973	GU737028	GU737183	GU736844	GU736749

	GENBANK ACCESSION NUMBERS				
	Mitochondrial			Nuclear	
Taxon	Cyt *b*	ND4	16S rRNA	RAG2	S7
Cichlini					
Cichla intermedia	GU736925	AY566788	GU737116	AY566752	GU736682
Cichla monoculus	DQ990686	GU736997	AF049017	-	-
Cichla orinocensis	AF370643	AY566786	GU737115	AY566751	GU736681
Cichla temensis	AF370644	AY566793	GU737114	AY566755	GU736680
Geophagini					
Acarichthys heckelii	AF370653	AY566768	GU737120	AY566733	GU736686
Acarichthys heckelii-Guyana	DQ990687	GU737000	GU737121	-	GU736687
Apisitogramma agassizi	-	AY566787	GU737122	AY566749	GU736688
Apistogramma hoignei	GU736927	AY566781	GU737123	AY566746	GU736689
Apistogramma iniridae	-	-	GU737124	-	GU736690
Apistogramma pucallpaensis	-	AY566770	GU737125	AY566735	GU736691
Biotodoma cupido	GU736928	AY566772	GU737126	AY566723	GU736692
Biotodoma wavrini	AF370657	AY566784	GU737127	AY566726	GU736693
Biotoecus dicentrarchus	GU736929	AY566792	GU737128	AY566754	GU736694
Crenicara punctulata	AF370655	-	GU737129	AY566742	GU736695
Crenicichla sp. ‘Orinoco lugubris’	GU736932	AY566785	GU737132	AY566750	GU736698
Crenicichla sp. ‘Orinoco wallaci’	GU736935	AY566790	GU737135	AY566753	GU736701
Crenicichla geayi	GU736930	AY566771	GU737130	AY566736	GU736696

	GENBANK ACCESSION NUMBERS				
	Mitochondrial			Nuclear	
Taxon	Cyt *b*	ND4	16S rRNA	RAG2	S7
Crenicichla lenticulata	GU736931	GU737001	GU737131	GU736813	GU736697
Crenicichla lugubris - Guyana	GU736933	GU737002	GU737133	GU736814	GU736699
Crenicichla multispinosa	GU736936	-	GU737136	GU736815	GU736702
Crenicichla sveni	GU736934	AY566779	GU737134	AY566743	GU736700
Crenicichla reticulata	GU736937	GU737003	GU737137	GU736816	GU736703
Dicrossus sp.	GU736938	AY566767	GU737138	AY566731	GU736704
Geophagus abalios	GU736939	AY566795	GU737139	AY566757	GU736705
Geophagus dicrozoster	GU736941	AY566794	GU737141	AY566756	GU736707
Geophagus grammepareius	GU736942	AY566796	GU737142	AY566724	GU736708
Geophagus harreri	GU736943	GU737004	GU737143	GU736817	GU736709
Geophagus sp. 'Cuyuni'	GU736940	-	GU737140	AY566727	GU736706
Geophagus sp. 'Takutu'	GU736945	GU737005	GU737145	GU736818	GU736711
Geophagus surinamensis	GU736944	AY566777	GU737144	AY566741	GU736710
Geophagus taeniopareius	GU736946	GU737006	GU737146	GU736819	GU736712
'Geophagus' brasiliensis	AF370659	AY566766	GU737148	AY566732	GU736713
'Geophagus' steindachneri	AF370660	AY566765	GU737147	AY566730	DQ119275
Guianacara sp. 'Takutu'	GU736947	GU737007	GU737149	GU736820	GU736714
Guianacara owroewefi	-	-	GU737152	-	-
Guianacara stergiosi - Caroni	GU736948	AY566762	GU737150	AY566725	GU736715

	GENBANK ACCESSION NUMBERS				
	Mitochondrial			Nuclear	
Taxon	Cyt *b*	ND4	16S rRNA	RAG2	S7
Guianacara stergiosi - Aro	GU736949	GU737008	GU737151	GU736821	GU736716
Gymnogeophagus balzanii	GU736950	-	GU737153	AY566739	GU736717
Gymnogeophagus rhabdotus	GU736951	AY566775	GU737154	AY566738	GU736718
Gymnogeophagus setequedas	GU736952	GU737009	GU737155	GU736822	GU736719
Mazarunia mazarunii	GU736960	GU737012	GU737165	GU736825	GU736729
Mazarunia sp. 1	GU736962	GU737014	GU737167	GU736827	GU736731
Mazarunia sp. 2	GU736961	GU737013	GU737166	GU736826	GU736730
Mikrogeophagus altispinosus	GU736953	AY566764	GU737156	AY566729	GU736720
Mikrogeophagus ramirezi	GU736954	AY566780	GU737157	AY566744	GU736721
Santanoperca daemon	GU736955	AY566791	GU737158	AY566758	GU736722
Santanoperca jurupari	AF370664	AY566783	GU737159	AY566745	GU736723
Santanoperca leucosticta	GU736956	GU737010	GU737160	GU736823	GU736724
Santanoperca mapiritensis	GU736957	AY566761	GU737161	AY566728	GU736725
Santanoperca pappaterra	GU736958	AY566773	GU737162	AY566759	GU736726
Taeniacara candidi	AF370665	AY566769	GU737163	AY566734	GU736727
Teleocichla aff. *proselytus*	GU736959	GU737011	GU737164	GU736824	GU736728
Heroini					
Amatitlania siquia	AY843376	GU737054	GU737208	GU736870	DQ119254
Amphilophus citrinellus	AY843348	GU737045	GU737199	GU736861	DQ119256

	GENBANK ACCESSION NUMBERS				
	Mitochondrial			Nuclear	
Taxon	Cyt *b*	ND4	16S rRNA	RAG2	S7
Amphilophus hoogabomorus	AY843433	GU737046	GU737200	GU736862	GU736765
Amphilophus labiatus	U88863	GU737047	GU737201	GU736863	GU736766
Archocentrus centrarchus	AF009931	GU737049	GU737203	GU736865	EF433033
Astatheros diquis	AF009945	GU737056	GU737210	GU736872	GU736773
Astatheros longimanus	GU736983	GU737057	GU737211	GU736873	GU736774
Astatheros macracanthus	DQ990695	GU737076	GU737230	GU736890	GU736785
Astatheros octofasciatus	AY843410	GU737062	GU737216	GU736877	GU736777
Astatheros robertsoni	AF145132	GU737058	GU737212	-	GU736775
Astatheros rostratus	AF009944	GU737059	GU737213	GU736874	-
Australoheros facetus	AY998666	GU737032	GU737186	GU736848	GU736752
Caquetaia krausii	AF009938	GU737033	GU737187	-	GU736753
Caquetaia umbrifera	AF009940	GU737070	GU737224	GU736885	EF433021
'Cichlasoma' festae	AY843351	GU737031	DQ119187	GU736847	EF433008
'Cichlasoma' calobrensis	AY843378	GU737063	GU737217	GU736878	GU736778
'Cichlasoma' grammodes	DQ990718	GU737068	GU737222	GU736883	EF433016
'Cichlasoma' lyonsi	AY843395	GU737048	GU737202	GU736864	GU736767
'Cichlasoma' salvini - Belize	AY050619	GU737065	GU737219	GU736880	DQ119258
'Cichlasoma' salvini - Mexico	-	GU737066	GU737220	GU736881	-
'Cichlasoma' trimaculatus	AY843407	GU737069	GU737223	GU736884	EF433018

	GENBANK ACCESSION NUMBERS				
	Mitochondrial			Nuclear	
Taxon	Cyt *b*	ND4	16S rRNA	RAG2	S7
'Cichlasoma' urophtahlmus	AY843427	-	GU737221	GU736882	EU620420
'Cichlasoma' wesseli	AY843384	GU737100	GU737253	GU736913	GU736801
'Cryptoheros' (Bussingius) myrnae	AF009927	GU737050	GU737204	GU736866	GU736768
'Cryptoheros' (Bussingius) nanoluteus	AY843398	GU737051	GU737205	GU736867	GU736769
'Cryptoheros' (Bussingius) sajica	AF009925	GU737052	GU737206	GU736868	GU736770
Cryptoheros (Cryptoheros) chetumalensis	AY843379	GU737053	GU737207	GU736869	GU736771
Cryptoheros (Cryptoheros) cutteri	GU736982	GU737055	GU737209	GU736871	GU736772
'Cryptoheros' (Panamius) panamensis	AY843354	GU737064	GU737218	GU736879	GU736779
Herichthys bartoni	DQ990721	GU737074	GU737228	-	GU736783
Herichthys carpintis	AY323999	GU737071	GU737225	GU736886	GU736780
Herichthys cyanoguttatus	AY323987	GU737072	GU737226	GU736887	GU736781
Herichthys labridens	AY323993	GU737073	GU737227	GU736888	GU736782
Herichthys pantostictus	AY323988	GU737077	GU737231	GU736891	GU736786
Herichthys steindachneri	AY324013	GU737078	GU737232	GU736892	GU736787
Herichthys tamasopoensis	AY324001	GU737075	GU737229	GU736889	GU36784
Heroina isonycterina	GU736981	GU737044	GU737198	GU736860	EF433044
Heros efasciatus	DQ010102	GU737034	GU737188	GU736849	DQ119276
Heros severus - Venezuela	GU736977	GU737036	GU737190	GU736851	GU736755
Heros severus - Nanay	GU736976	GU737035	GU737189	GU736850	GU736754

	GENBANK ACCESSION NUMBERS				
	Mitochondrial			Nuclear	
Taxon	Cyt *b*	ND4	16S rRNA	RAG2	S7
Heros sp. 'common'	GU736978	-	GU737191	GU736852	GU736756
Herotilapia multispinosa	-	GU737079	GU737233	GU736893	GU736788
Hoplarchus psittacus	AF370673	AY566789	GU737192	AY566760	GU736757
Hypselecara coryphaenoides	AF370674	GU737037	GU737193	GU736853	GU736758
Hypsophrys nematopus	AF009928	GU737081	GU737235	GU736895	GU736789
Hypsophrys unimaculatus	AF009930	GU737080	GU737234	GU736894	DQ119260
Mesonauta egregius	GU736979	AY566782	GU737194	AY566748	GU736759
Mesonauta insignis	AF370675	GU737038	AF045859	GU736854	GU736760
Nandopsis haitiensis	AY843388	GU737086	GU737240	GU736900	DQ119271
Nandopsis tetracanthus	AY998669	GU737087	GU737241	GU736901	DQ119270
Parachromis dovii	DQ990701	GU737082	GU737236	GU736896	GU736790
Parachromis friedrichsthalii - Belize	GU736985	GU737083	GU737237	GU736897	GU736791
Parachromis loisellei	AY843366	GU737085	GU737239	GU736899	GU736793
Parachromis managuensis	AY843356	GU737084	GU737238	GU736898	GU736792
Paraneetroplus bifasciatus - Mexico	GU736989	GU737101	GU737254	GU736914	GU736802
Paraneetroplus bulleri	AY324004	-	-	-	EU620423
Paraneetroplus guttulatus	AY324025	GU737103	GU737258	GU736918	GU736806
Paraneetroplus maculicauda - Belize	GU736991	GU737105	GU737260	-	-
Paraneetroplus maculicauda - Costa Rica	-	GU737104	GU737259	GU736919	GU736807

	GENBANK ACCESSION NUMBERS				
	Mitochondrial			Nuclear	
Taxon	Cyt *b*	ND4	16S rRNA	RAG2	S7
Paraneetroplus melanurus - Belize	AY843420	GU737106	GU737261	GU736920	GU736808
Paraneetroplus synspilus - Mexico	AY050625	GU737108	GU737263	GU736922	GU736809
Petenia splendida	AY843423	GU737088	GU737242	GU736902	DQ119264
Pterophyllum scalare (Scalare Genbank)	AF370676	GU737039	AY662732	GU736855	GU736761
Symphysodon aequifasciatus	AF370677	GU737040	GU737195	GU736856	GU736762
Symphysodon discus	AY840119	GU737041	GU737196	GU736857	GU736763
Theraps bocourti	GU736984	GU737061	GU737215	GU736876	GU736776
Theraps godmanni - Belize	AY843428	-	GU737255	GU736915	GU736803
Theraps intermedius - Belize	AY843408	-	GU737256	GU736916	GU736804
Theraps intermedius - Mexico	GU736990	GU737102	GU737257	GU736917	GU736805
Theraps irregularis	DQ494383	GU737098	GU737251	GU736911	GU736799
Theraps lentiginosus	AY843409	GU737099	GU737252	GU736912	GU736800
Theraps nourissati	EF436465	GU737060	GU737214	GU736875	EF433048
Theraps ufermanni	GU736992	GU737107	GU737262	GU736921	-
Thorichthys ellioti	AY324011	GU737092	GU737246	GU736905	GU736794
Thorichthys meeki	AY843426	GU737091	GU737245	GU736904	-
Thorichthys affinis	GU736987	GU737094	GU737248	GU736907	GU736796
Thorichthys aureus	U88859	GU737089	GU737243	GU736903	DQ119265
Thorichthys helleri	AY324022	GU737093	GU737247	GU736906	GU736795

	GENBANK ACCESSION NUMBERS				
	Mitochondrial			Nuclear	
Taxon	Cyt *b*	ND4	16S rRNA	RAG2	S7
Thorichthys meeki - Belize	GU736986	GU737090	GU737244	-	-
Thorichthys pasionis	DQ494385	GU737095	GU737249	GU736908	GU736797
Tomocichla asfraci	GU736988	GU737096	AY662735	GU736909	GU736798
Tomocichla tuba	AF009941	GU737097	GU737250	GU736910	JX437629
Uaru amphiacanthoides	AF370678	GU737042	DQ119191	GU736858	DQ119278
Uaru fernandezyepezi	GU736980	GU737043	GU737197	GU736859	GU736764
Retroculini					
Retroculus sp.	AF370640	AY566774	GU737119	AY566737	GU736685

ADDENDUM C

A summary of the 180 taxa included in the time-calibrated BEAST phylogenetic analysis. The four taxon sets and their inclusive species, as specified in the BEAST XML input file, for the BEAST analysis performed on the aligned Cyt *b* sequence matrix of 1,119 bp (Friedman et al., 2013; López-Fernández et al., 2010).

TAXON SET	TAXA INCLUDED
Cichlinae	*Acarichthys heckelii*
	Acarichthys heckelii Guyana
	Acaronia nassa
	Acaronia vultuosa
	Aequidens diadema
	Aequidens tetramerus
	Aequidens tetramerus Xingu
	Amatitlania siquia
	Amphilophus citrinellus
	Amphilophus hoogabomorus
	Amphilophus labiatus
	Andinoacara coeruleopunctatus
	Andinoacara pulcher
	Andinoacara rivulatus
	Apistogramma hoignei
	Archocentrus centrarchus
	Astatheros diquis
	Astatheros longimanus
	Astatheros macracanthus
	Astatheros octofasciatus
	Astatheros robertsoni
	Astatheros rostratus
	Astronotus sp.
	Australoheros fascetus
	Biotodoma cupido
	Biotodoma wavrini
	Biotoecus dicentrarchus
	Bujurquina aff apoparuana

TAXON SET	TAXA INCLUDED
	Bujurquina aff hophrys
	Caquetaia krausii
	Caquetaia umbrifera
	Chaetobranchopsis orbicularis
	Chaetobranchus flavescens
	Cichla intermedia
	Cichla monoculus
	Cichla orinocensis
	Cichla temensis
	Cichlasoma amazonarum
	Cichlasoma calobrensis
	Cichlasoma dimerus
	Cichlasoma festae
	Cichlasoma grammodes
	Cichlasoma lyonsi
	Cichlasoma orinocense
	Cichlasoma salvini Belize
	Cichlasoma trimaculatus
	Cichlasoma urophthalmus
	Cichlasoma wesseli
	Cleithracara maronii
	Crenicara punctulata
	Crenicichla geayi
	Crenicichla lenticulata
	Crenicichla lugubris Guyana
	Crenicichla multispinosa
	Crenicichla reticulata
	Crenicichla sp Orinoco lugubris
	Crenicichla sp Orinoco wallaci
	Crenicichla sveni
	Cryptoheros chetumalensis
	Crypthoheros cutteri

TAXON SET	TAXA INCLUDED
Cichlinae	*Crypthoheros myrnae*
	Crypthoheros nanoluteus
	Cryptoheros panamensis
	Cryptoheros sajica
	Dicrossus sp
	Geophagus abalios
	Geophagus brasiliensis
	Geophagus dicrozoster
	Geophagus grammepareius
	Geophagus harreri
	Geophagus sp Cuyuni
	Geophagus sp Takutu
	Geophagus steindachneri
	Geophagus surinamensis
	Geophagus taeniopareius
	Guianacara sp Takutu
	Guianacara stergiosi Aro
	Guianacara stergiosi Caroni
	Gymnogeophagus balzanii
	Gymnogeophagus rhabdotus
	Gymnogeophagus setequedas
	Herichthys bartoni
	Herichthys carpintis
	Herichthys cyanoguttatus
	Herichthys labridens
	Herichthys pantostictus
	Herichthys steindachneri
	Herichthys tamasopoensis
	Heroina isonycterina
	Heros efasciatus
	Heros severus Nanay
	Heros severus Venezuela

TAXON SET	TAXA INCLUDED
Cichlinae	*Heros sp common*
	Hoplarchus psittacus
	Hypselecara corypaenoides
	Hypsophrys nematopus
	Hypsophrys unimaculatus
	Krobia potaroensis
	Krobia sp Xingu orange spot
	Laetacara dorsigera
	Laetacara thayeri
	Mazarunia mazarunii
	Mazarunia sp 1
	Mazarunia sp 2
	Mesonauta egregius
	Mesonauta insignis
	Mikrogeophagus altispinosus
	Mikrogeophagus ramirezi
	Nandopsis haitensis
	Nandopsis tetracanthus
	Nannacara taenia
	Parachromis dovii
	Parachromis friedrichsthalii Belize
	Parachromis loisellei
	Parachromis managuensis
	Paraneetroplus bifasciatus Mexico
	Paraneetroplus bulleri
	Paraneetroplus guttulatus
	Paraneetroplus maculicauda Belize
	Paraneetroplus melanurus Belize
	Paraneetroplus synspilus Mexico
	Petenia splendida
	Pt. leopoldi C
	Pt. leopoldi D

TAXON SET	TAXA INCLUDED
Cichlinae	*Pt. altum* Rio Negro 1
	Pt. altum Rio Negro 2
	Pt. altum Rio Atabapo 1
	Pt. altum Rio Atabapo 2
	Pt. altum Rio Inirida 1
	Pt. altum Rio Inirida 2
	Pt. altum Rio Ventuari 3
	Pt. altum Rio Ventuari 4
	Pt. scalare Genbank
	Pt. scalare Manacapuru A
	Pt. scalare Manacapuru B
	Pt. scalare Germany B
	Pt. scalare Germany D
	Pt. scalare Guyana A
	Pt. scalare Guyana B
	Pt. scalare Peru 2A
	Pt. scalare Peru 2B
	Pt. scalare Xingu B
	Pt. scalare Xingu C
	Retroculus sp
	Santanoperca daemon
	Santanoperca jurupari
	Santanoperca leucosticta
	Santanoperca mapiritensis
	Santanoperca pappaterra
	Symphysodon aequifasciatus
	Symphysodon discus
	Taeniacara candidi
	Tahuantinsuyoa macantzatza
	Teleocichla aff proselytus
	Theraps bocourti
	Theraps godmani Belize

TAXON SET	TAXA INCLUDED
Cichlinae	*Theraps intermedius Belize*
	Theraps intermedius Mexico
	Theraps irregularis
	Theraps lentiginosus
	Theraps nourissati
	Theraps ufermanni
	Thorichthys affinis
	Thorichthys aureus
	Thorichthys ellioti
	Thorichthys helleri
	Thorichthys meeki
	Thorichthys meeki Belize
	Thorichthys pasionis
	Tomocichla asfraci
	Tomocichla tuba
	Uaru amphiacanthoides
	Uaru fernandezyepezi
Etroplinae	*Etroplus maculatus*
	Paretroplus polyactis
Ptychochrominae	*Paratilapia polleni*
	Ptychochromis oligocantus
Pseudocrenilabrinae	*Chromidotilapia guntheri*
	Hemichromis fasciatus
	Heterochromis multidens