MOLECULAR SYSTEMATICS OF THE SOUTH AFRICAN GOBIID FISHES OF THE GENUS *GLOSSOGOBIUS* GILL, 1862 (TELEOSTEI: GOBIIDAE)

by

PHOLOSHI ABRAM MAAKE

RESEARCH DISSERTATION

Submitted in fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in

ZOOLOGY

in the

FACULTY OF SCIENCE & AGRICULTURE

(School of Molecular and Life Sciences)

at the

UNIVERSITY OF LIMPOPO

SUPERVISOR: Prof SM Dippenaar

CO-SUPERVISOR: Dr M Mwale

2009

Declaration

I declare that the dissertation hereby submitted to the University of Limpopo, for the degree of Master of Science in Zoology has not previously been submitted by me for a degree at this or any other university; that it is my own work in design and in execution, and that all material contained therein has been duly acknowledged.

_____ _____

P.A. Maake (Mr) **Date**

ABSTRACT

The genus, *Glossogobius* Gill, 1862 belongs to the family Gobiidae, a large fish family represented by several thousand species. In southern African there are about 51 genera comprising about 107 species. It has been suggested that gobies underwent an extensive ecological radiation, with species diversifying into marine, brackish and freshwater habitats. The genus *Glossogobius* has about 22 species with an Indo-west Pacific distribution. The overall aim of the current study was to examine the systematics of *Glossogobius* in southern Africa using molecular markers in an attempt to clarify the geographic distribution and species status.

A total of 131 individuals of *G. callidus* were collected from rivers and estuaries throughout the species' distribution in two southern African biogeographic regions (subtropical and warm-temperate) using six-meter seine nets, while individuals of *G. tenuiformis* were only collected in KwaZulu-Natal region. Additional specimens and taxa were downloaded from Genbank and also loaned from the SAIAB DNA collection and the Art Gallery and Museum of the Northern Territory (NTM), Darwin Australia. This study employed mtDNA cytochrome *b* (cyt *b*) and the nuclear S7 ribosomal protein gene intron 1 (S7 intron 1) as genetic markers. In determining the phylogeographic structure of the river goby general methods such as AMOVA, demography analyses, network of haplotypes and phylogenetic tree reconstruction were employed. Evolutionary relationships within the genus were resolved by employing parsimony (MP), Bayesian Inference (BI) and Maximum likelihood (ML) tree reconstruction analyses. Both MP and ML were computed in PAUP* while BI was computed using Mr Bayes.

Initially, a large-scale assessment of the genetic structure and diversity of *G. callidus* was undertaken. This species has a wide distribution in most estuaries and coastal river systems along the east coast of southern Africa, and in inland freshwater systems from the north-eastern parts. Overall, patterns of haplotype relatedness, high haplotypic diversity related to low nucleotide diversity, and mismatch distribution analysis supported a population expansion hypothesis with continued restricted dispersal and gene flow. Two major historically isolated lineages (5% divergence, cyt *b*) were identified within this species, one comprising subtropical samples (KwaZulu-Natal lineage), and the other including the majority of the warm temperate localities and a few subtropical localities (EC, Eastern Cape lineage). *Glossogobius callidus* individuals collected from riverine localities within the EC lineage were geographically structured and support the general hypothesis of structure among freshwater species while the estuarine individuals were less structured. Although, the

individuals from these two lineages were morphologically inseparable, the high genetic separation between the two conforms to the reports that *G. callidus* is a species complex. The results further revealed some evolutionary significant units (Nkanini/Kosi bay lineage, and Mozambique) in the KZN lineage, KwaZulu-Natal. However, there were other identified localities with restricted gene flow and some degree of reproductive isolation that were not evolutionary significant. Thus, the results indicate that *G. callidus* is a species with relatively strong population and phylogeographic structure in southern Africa.

The validity of the species of the genus *Glossogobius* in southern Africa (*G. callidus* and *G. giurus* with its synonymy *G. tenuiformis*) was investigated by estimating the relationships among the *Glossogobius* taxa in southern Africa, using the same molecular markers. Based on a model selected under AIC in Modeltest, ML analysis of the dataset also recovered a phylogram which was similar to MP and BI in recognising the major clades in many respects. All analyses recovered monophyly for all southern African species of the genus. Although the cyt *b* analyses included *Glossogobius* species from Australia, there was no evidence of a phylogeographic break between the two continents. However, it appears that each southern African *Glossogobius* species has a sister species occurring in Australia according to the cyt *b* dataset. *Glossogobius aureus* was revealed as the probable sister species to *G. tenuiformis* while the African *G. giuris* group was monophyletic with the Australian *G. giuris*. The two *G. callidus* groups that appear to represent two unique species (sp. 1 and sp. 2) were also sister taxa to the undescribed *Glossogobius* sp. from Australia. It could be postulated that this inter-continental species relationships were a result of either allopatric or sympatric speciation events, to the ancestral taxa of each of these three sister groups which in turn split up during the Gondwana breakup. The inclusion of *G. celebius* (a freshwater and brackish species with an Asian and Oceania distribution) in the S7 intron dataset did not affect the monophyly of southern African species. However, this species appears to be more basal in the genus than the southern African species. The results further indicated that *G. tenuiformis* (previously a synonym of *G. giuris*) forms a separate clade that is distinct from the *G. giuris* species group. Although *G. tenuiformis* and *G. giuris* formed a monophyletic group in all analyses involving the S7 intron 1, individuals from each species were monophyletic and strongly supported. The scanning electron micrographs of the cheek papillae of the two species further revealed morphological differences. The congruence between both morphology and molecular data indicated that *G. tenuiformis* is unequivocally distinct from *G. giuris* and it is therefore suggested that the taxonomic status of the *G. giuris* group be revised.

In conclusion, both genetic markers were successfully used in the current study in identifying the phylogeographic structure of the river goby as well as determining the phylogenetic relationships of the southern African flat head *Glossogobius* species. The mtDNA data provided good clarity in the genetic structure of *G. callidus* as well as in solving the deeper phylogenetic relationships of the genus. Although, the S7 intron 1 produced strong support in delineating deeper relationships, i.e. those between species, this gene was not as effective for intraspecific relationships. The phylogenetic results presented in this study show that the diversity of the genus in southern Africa was underestimated. *Glossogobius callidus* clade B (sp. 2) and *G. tenuiformis* emerged as distinct taxa such that the number of the species in our region has doubled. New phylogenetic information can now be added to morphological knowledge to produce a more accurate taxonomic framework for *Glossogobius* in our region and elsewhere.

ACKNOWLEDGEMENTS

Firstly I would like to thank all members of the South African Institute for Aquatic Biodiversity (SAIAB) for accepting me into their home and hearts during all my time and stay at the institute. From the first day I arrived at the institute, all your smiles and love made me feel at home "away from home". Also to all my friends your friendship during these years is invaluable.

This research and degree was sponsored by funds from the National Research Foundation (NRF), Ernst and Ethel Eriksen Trust and SAIAB. I also thank the University of Limpopo postgraduate desk for providing tuition waver funds during academic year of this degree.

I thank my supervisors Dr. Monica Mwale and Prof. Susan Dippenaar, and Mr. Ofer Gon (member of my supervisory committee) for their time, encouragement and invaluable advice throughout my research career thus far. Thanks for walking the extra mile and for giving me so many opportunities and financial support to do field surveys, extra lab work and to attend conferences and workshops. Without your support and mentorship this thesis would not have been possible.

To Drs. Gavin Gouws and Ernst Swartz, I pledge my special thanks for all the times you have put on helping me with different types of analyses, discussions and all the friendly talks we had all the way. Without all of this the journey I took towards my writing would have been rather "mission impossible".

I would also like to thank Ms Poogendri Reddy who took time to guide me through the methods and procedures used in the genetics laboratory of SAIAB. "If you didn't take your time to put labels on all the reagents and apparatus in the laboratory, I sure know that I would have caused chaos". It would be a sin if I would not thank my laboratory mates (Caswel Mavimbela, Mpho Ramoejane and Tshoanelo Moloi) for the friendly working atmosphere and their time for helping when things seemed to go wrong for me. I also like to extent my thanks to Mpho Ramoejane for providing me with extra DNA materials from localities that I was not able to sample myself. I thank Kate Munik who reserved a seat for me when she and her supervisor, Tony Booth, went on a sampling trip that I could not manage by myself. I thank Paul Cowley for generously providing me with additional DNA material from Kleinemonde area.

To the students and staff in the Department of Ichthyology and Fisheries Science (DIFS), Rhodes University, I would like to thank them for their precious time that they used to attend my calls for presentation and their friendship during the years of my stay in Grahamstown. To Mandilakhe Mkalipi and Bafo Kongqobe, I thank you very much for all the times you took to help me know my ways in the fish collection facility of SAIAB and availing yourself when I needed to loan materials from the facility.

Special thanks also goes to SAIAB DNA banking biomaterial officer, Unathi Lwana for providing me with the extra needed DNA material from other parts of the southern African region. I would like to acknowledge the generosity of Dr Helen Larson of the Art Gallery and Museum of the Northern Territory (NTM), Darwin, Australia, for providing the additional taxa used in the present study.

Finally, I would like to thank my parents, Solly and Neo Maake, and my two brothers, Carlvien and Sebo, for all their love, and their constant emotional, moral, and financial support during all stages of my life. I am what I am today because you loved me unconditionally…..

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

GENERAL INTRODUCTION

1. LITERATURE REVIEW

Fish dispersal within and among terrestrial aquatic systems (e.g. rivers and estuaries) is restricted by their physiological and other biological attributes as well as the past and present connectivity of the systems (Swartz *et al*. 2007). Climatic and geological changes underlie the geomorphological history of drainage systems and define the resultant connectivity within and between the fish species in these systems. Following historical climate shifts and geological changes, species range expansion and contraction can result in populations that are isolated by spatial and temporal barriers over a considerable area (Hewitt 2000; Swartz *et al.* 2007). Genetic divergence may occur over time once barriers between drainage systems become established and disrupt gene flow between the adjacent populations (Stepien and Faber 1998; Benzie 1999a). This disruption of gene flow has been found to play a major role in shaping the history of species and populations in most phylogeographic studies of freshwater fishes (Skelton 1980a, b). For example Swartz *et al*. (2007) reported that in *Pseudobarbus afer* and *P. phlegethon* this is primarily because of the restricted distribution and the strong association of the rivers with the geological events in freshwater fishes. Consequently, adjacent populations may not be homogenous in genetic makeup between drainage systems due to disruption of gene flow (Benzie 1999b; Swartz *et al*. 2007).

Two other geological phenomena can be generally applied to explain the historical connections among drainages: the first is river capture, which is often the case when fishes from different river systems share a recent history of gene flow after contact (Skelton 1980a; Swartz *et al*. 2007). A typical example of river capture according to Skelton (1980a) is when a river breaks through a barrier such as a mountain range and captures the headwaters of another drainage system. River capture can influence the connectivity and dispersal of fish species to new drainages when contacts between systems have been established (Swartz *et al*. 2007). The second phenomenon can occur due to river confluence after sea level changes as adjacent drainages or estuaries merge. Sea level changes could have been responsible for many cycles of establishment and disruptions of river connections during glacial and interglacial times (Skelton 1980b; Blum and Tornquist 2000; Swartz *et al*. 2007). Based on a biogeographical historical perspective of southern Africa, the dispersal and

isolation of fish into many of the smaller coastal and terrestrial systems has taken place during periods of low sea level (Skelton 1980b).

An extensive body of work in molecular genetic analyses on fishes which has focused on and answered the questions about the genetic makeup of species (phylogeny and taxonomy of species) and populations (phylogeography and population structure) affected by the above phenomena is available in the literature (Avise *et al*. 1987; Moritz *et al*. 1987; Davies-Coleman *et al*. 2000; Schlotterer 2001; Harada *et al*. 2002; Thacker 2003; Fujita *et al*. 2004; Qi *et al*. 2007). The geographic distribution of genetic variation within and among taxa has been used to provide information on these historical, contemporary demographic and evolutionary processes among fish species (Avise *et al*. 1987; Avise 2000). It is thus hypothesised and can be expected that the phylogeographic structure of freshwater fishes could to some extent reflect the inability of species less able to disperse, to cross the biogeographic barriers explained above, as well as the effect of past river capture or sea level changes. This may result in genetic divergence of isolated units of such species due to low levels of gene flow (Tzeng *et al*. 2006). There are several South African species that have been suggested to exhibit high levels of genetic structuring as a consequence of these factors. Swartz *et al*. (2007) reported a significant genetic structuring among the freshwater fiery redfins in the Eastern Cape. They suggested that the biogeographic history of the South African coast and the river capture in the Eastern Cape are the major cause of the species phylogeographic structure.

Although freshwater fishes are generally hypothesised to show high levels of structuring, most estuarine fishes are expected to possess little to no genetic variation across their distribution (Grant and Bowen 1998; Oosthuizen *et al*. 2004). This originates from the assumption that tides, waves and currents provide an efficient dispersal mechanism for eggs, larvae, and adults during the open phase of the estuary mouth (Grant and Bowen 1998; Whitfield 1998). However, this assumption, if untested, may have serious implications on current research strategies and the provision of useful information on the biodiversity of threatened environments and species. Recently, Oosthuizen *et al*. (2004) used the mitochondrial DNA (mtDNA) cytochrome oxidase subunit III (COIII) gene to study a cephalopod, *Octopus vulgaris* Cuvier, 1797, on the coast of South Africa. There was no sequence variation within and between all the samples from the examined localities supporting a single population for this estuarine species. Neethling *et al*. (2008) also found that there was high level of gene flow between populations of the South African endemic coastal goby, *Caffrogobius caffer* Günther, 1874, which indicated that waves and currents do have an effect on migration within this

region and therefore encourage some gene flow between populations. However, in their population genetics study of *Hippocampus capensis* Boulenger, 1900, Teske *et al*. (2003) discovered that all three known populations of the species are genetically distinct management units. They thus pointed out that the single population hypothesis for estuarine species was incorrect for this South African endemic and endangered estuarine seahorse.

Although some work has been done on the phylogeography of gobiid fishes testing the above mentioned two hypotheses, e.g. studies on *Eucyclogobius newberryi* Gill, 1862 (Dawson *et al.* 2001) and *Pomatoschistus minutus* Pallas, 1770 (Stefanni and Thorley 2003), only one population genetic study has been done on species of *Glossogobius* Gill, 1862 in South Africa (Engelbrecht and Mulder 1998). This particular study on the genus was conducted using protein electrophoreses. Engelbrecht and Mulder (1998) compared the extent of genetic variation and differentiation within and between two populations of *G. callidus* Smith, 1937, one estuarine (from Nhlambe estuary in KwaZulu-Natal Province) and the other riverine (from the Olifants River system in the Limpopo Province). Although their results revealed considerable variation, these authors could not conclude whether the populations in the two sampled environments were reproductively isolated and/or representing two genetically distinct species of the genus due to poor distribution representation. For the purpose of conservation, the assessment of the phylogeography of a species' entire distribution and the identification of genetically divergent areas are of fundamental importance for proper interpretation of results (Moritz 1995; Stefanni and Thorley 2003).

Examples of studies highlighting the importance of examining genetic variation in conservation of fishes includes the identification of an evolutionary significant unit (ESU) in the sand goby (*Pomatocshistus minutes* Pallas, 1770) of the Adriatic Sea in the Mediterranean (Stefanni and Thorley 2003). Von der Heyden *et al.* (2007) used genetic variation to unravel the evolutionary history of sympatric species (*Merluccius capensis* Castelnau, 1861 and *Merluccius paradoxus* Franca, 1960), as well as addressing management issues within regions where commercially valuable fish stocks are shared between nations. Teske *et al*. (2003) also demonstrated the important value of such studies in assessing the conservation status of natural populations and for the proclamation of protected areas as in the case for the Knysna seahorse. Thus molecular methods can provide an objective means for setting conservation priorities and may assist in the focusing of conservation efforts on specific populations that are in need of recognition and recovery (Stefanni and Thorley 2003). Altukhov (2003) asked the question "why genetic stability and structure matter so much?." He argued that this is because the sustainable use of natural resources, including genetic resources of

populations is critically dependant on the maintenance of their stability and that the preservation of well-adapted genetic characteristics from one generation to the next is essential for this stability.

The assessment of the genetic variation and phylogeography of a species' entire distribution is important for genera such as *Glossogobius* where wide variation occurs in the diagnostic morphological characters that are commonly used for species delineation (Akihito *et al.* 2000; Thacker and Cole 2002; Thacker 2003). The observed morphological differences within species of *Glossogobius* have raised questions among goby systematists about the taxonomic status of its species and their composition (Hoese 1986; Akihito and Meguro 1975). Although several systematic studies on the phylogeny and taxonomy of gobies have been undertaken, it has been difficult or impossible for systematists to accurately discern species boundaries due to this high morphological variability (Harrison 1989; Hoese and Gill 1993). This problem within the species of the genus *Glossogobius* has resulted in several synonymies among known species. In South Africa, identification of species of *Glossogobius* has also been plagued by considerable variability in diagnostic morphological features, resulting in confusion and often misidentification of species (Hoese and Winterbottom 1979; Hoese 1986).

These taxonomic problems have also been fuelled by the prevalence of reductive evolution among the species of this genus that have made identification of informative morphological characters particularly difficult (Smith 1960; Hoese 1986; Thacker 2003; Thacker and Hardman 2005). The taxonomic distinctiveness of species has important implications for the assignment of conservation priorities of biodiversity and species protection for survival (Hillis and Moritz 1990). Because of the importance of species identification and conservation, most gobies have increasingly been studied to answer the taxonomic, phylogenetic and evolutionary questions pertinent to relationships of and recognition of species, commonly using molecular techniques and data (Akihito *et al*. 2000; Dawson *et al*. 2001; Harada *et al*. 2002; Thacker and Cole 2002; Thacker 2003). These studies have generally indicated that morphological analyses alone are inadequate for revealing the presence of lineages that have diverged thousands to millions of years ago (recent and past) among these species in cases where diagnostic morphological characters where lacking (Colborn *et al*. 2001; Hou *et al*. 2007). However, studies using morphology are important in tracing the evolution of any number of specific physical traits and are essential to the field of taxonomy (Harrison 1989; Thacker 2003).

Therefore, the extensive regional distribution of *Glossogobius* in southern Africa and its within species morphological variation with a number of synonymies makes this genus a good candidate for

a phylogenetic analysis using different genetic markers, such as the mtDNA cytochrome *b* and nuclear DNA introns. Molecular analyses using these two genes have been useful in studying the systematics of several fish genera (Lavoue *et al*. 2003) and could be useful in understanding the systematics of the genus *Glossogobius* in southern Africa. The determination of the phylogeographic structure of *G. callidus,* a species that is generally considered to be a species complex could assist in testing the validity of the hypothesis of low genetic structure among estuarine populations as well as strong genetic structure in riverine species. The study of the genetic structure of species in estuarine and riverine habitats is even more important and urgent when one considers the increasing coastal development, as well as the threat of invasive species. These changes put species in these environments under pressure such that their fish populations may be adversely affected. Understanding their phylogeographic history therefore would be useful in conservation plans and setting up management procedures in cases where genetic diversity and unique evolutionary significant lineages are under threat.

1.1. South African biogeographic regions and the study areas

The South African coastline is governed by a range of climatic and oceanic conditions (see, Whitfield 2000). The coastline stretches for about 3400 km, from the Orange River mouth on the west coast to Kosi Bay on the east coast (Teske *et al*. 2006). The South African coastline has approximately 370 freshwater outlets to the sea ranging from small, temporarily open coastal streams or rivers to large permanently open tidal estuaries (Harrison 2004). Based on ichthyofaunal data, environment and geomorphological classification, three biogeographic regions that characterise this coastline environment have been identified and delineated (Whitfield 1998; Harrison 2004). These are the cool-temperate region, from the Orange River estuary to Cape Point, the warm-temperate region from Cape Agulhas to and including the Mbashe estuary, and the subtropical region from the Mbashe to Kosi Bay (Figure 1.1).

The warm-temperate and subtropical regions are influenced by one major ocean current, namely the Agulhas current. Being tropical in origin, the waters of the Agulhas current are warm but become cooler with the decline in coastal sea temperatures as the current flows further south and moves offshore (Harrison 2004). Towards northern KwaZulu-Natal, the Agulhas current has been observed to flow closer inshore due to the narrower continental shelf and gentler continental slope, forming an elongated system of eddies in this area (Payne and Crawford 1989). Here the inshore water is transported in the opposite direction to that of the main stream of the Agulhas current, while the current on the shelf edge forms a formidable barrier to the free exchange of water and biota with the

open ocean (Lutjeharms 2005). As these eddies and filaments formed by the current move slowly towards offshore further south of KwaZulu-Natal, the current flashes parcels of the water onto the shelf (Payne and Crawford 1989). These eddies could also encourage gene flow and result in homogeneous populations in this region. Whitfield (1990) mentioned that the inshore currents along this subtropical coastal region could also retain eggs, embryo, and fish larvae thereby restricting dispersal. If, however, spawning takes place in offshore waters along the coastline, eggs, embryo, and larvae are rapidly transported southwards in the fast flowing Agulhas current.

The Agulhas current system thus has a major effect on the unpredictable estuarine environments, which are driven by the variables of water quantity and tidal regimes (Whitfield 1998). Whitfield (1998) described an estuary simply as an area where a river meets a sea and indicated that these areas are highly variable and largely affected by the above mentioned factors. Fishes and other aquatic organisms residing in these systems respond to these changing conditions on temporal and spatial scales that vary from one biogeographic region to another as they are influenced by the Agulhas current (Van der Elst 1988; Whitfield 1998). Like estuaries, South African rivers are also known to be very dynamic, constantly changing because of the landscape and the seasonal rainfall (Skelton 1980a; Skelton 2001; Swartz *et al*. 2007). River systems have been defined as a combination of small headwater streams that begin in the mountains before eventually combining to form a larger river entering a lake or an estuary (Skelton 2001). These rivers are generally separated by the physical barriers which can also permanently isolate their residing populations (Skelton 1980a). Species such as members of *Glossogobius* prefer the middle reaches of the river where the bed widens, flow decreases and vegetation, such as reeds, become part of the habitat, making it a suitable habitat for these species (Skelton 2001). Unfortunately, middle reaches are areas where most southern African rivers are also exposed to human impact i.e., many weirs have been built and water is extracted for agricultural purposes. These interferences with river systems may impact on the survival of fish species, including the small sized *Glossogobius* species, by affecting water levels and movement patterns of the fishes. Understanding the factors affecting the structuring and panmixia of a species is particularly critical for goby species such as *G. callidus* which is known to inhabit riverine and estuarine systems that are now under considerable and increasing pressure from this development and from invasive species.

1.2. Studied taxa

The genus, *Glossogobius*, belongs to the family Gobiidae, which is a large family represented by several thousand species. Although most goby species occur in the tropical Indo-west Pacific, they are also abundant in the temperate waters of both the northern and southern hemispheres (Hoese 1986). There are about 51 genera comprising of about 107 species in the southern African region alone (Hoese 1986). Gobies are generally small, well camouflaged fishes that are found mainly in tidal pools, sea grass beds, on coral, coral rubble, rocky reefs and on sandy and muddy bottoms mostly in marine and estuarine environments (Heemstra and Heemstra 2004). They play an important role in the ecology of these habitats as they are abundant and form a major component of the fish biomass and the food web. Because of this abundance and widespread distribution, gobies are good candidate taxa for studying the history and evolution of these systems (Penzo *et al*. 1998).

It has been suggested that gobies underwent an extensive ecological radiation, with species diversifying into marine, brackish, and freshwater habitats (Penzo *et al*. 1998; Thacker 2003). The relationships between freshwater and marine gobies represent a long-standing unresolved phylogenetic question on the origin of this group. The family is presently more abundant in the marine system suggesting that this may be the major area of diversification before an invasion of freshwater habitats (Penzo *et al*. 1998). The difficulty in identifying morphological characters suitable for unambiguously establishing any goby phylogeny raises the question of whether the adaptation to freshwater habitats evolved only once or several times independently (van der Elst 1988; Penzo *et al*. 1998; Akihito *et al*. 2000; Thacker 2003). This question is pertinent for the widely distributed goby genera such as *Glossogobius* that have species that are distributed in both riverine and estuarine environments.

The genus *Glossogobius* has about 22 species (Table 1.1) with an Indo-west Pacific distribution (Hoese 1986; Nelson 2006). In southern Africa, the contemporary distribution of the genus encompasses a large range of temperate and tropical coastal habitats with a few riverine species distributions (Whitfield 1998). Similar to the situation in bays and estuaries of temperate Australia (Neira *et al*. 1992; Neira and Potter 1994), *Glossogobius* larvae often constitute the most abundant fish in plankton surveys in both temperate and subtropical estuaries in South Africa (Harris and Cyrus 2000; Strydom *et al*. 2003). Species of the genus *Glossogobius* are therefore interesting with regards to evolution because they have adapted to freshwater, estuarine and marine habitats being common benthic residents in these systems.

Although the major distinguishing morphological characteristics of *Glossogobius* species such as the depressed head with relatively long snouts, large mouths and a bi-lobed tongue are useful at the generic level, there are very few interspecific characters. The systematics of the genus is therefore questionable as the taxonomy of species is still in question. While two *Glossogobius* species, namely *Glossogobius callidus* Smith, 1937, and *Glossogobius giuris* Hamilton, 1822 have presently been recognised from South African waters, the genus has been plagued by a considerable number of synonymies (Akihito and Meguro 1975; Hoese 1986; Greenwood 1994). Hoese and Winterbottom (1979) reported that Smith,1936 described *Gobius gulosus* from South Africa which he (Smith) subsequently realised that it was a primary homonym of *Microgobius* (then *Gobius*) *gulosus* Girard, 1858. This species was then later renamed as *Gobius callidus* (Smith, 1937). Although Hoese and Winterbottom (1979) only recognised two common species (*G. callidus* and *G. giuris*) of the genus in South Africa, it is probable that a third species, *G. tenuiformis,* described by Fowler in 1934 is a valid species (unpublished work; Helen Larson, *pers. comm.*; Art Gallery and Museum of the Northern Territory). *Glossogobius tenuiformis* was not included in the study of Hoese (1986) as it was considered to be a junior synonym of *G. giuris.*Therefore, to date, *G. giuris* and *G. callidus* are the only two recognised species in southern Africa. *Glossogobius callidus* is differentiated from *G. giuris* as being a smaller-sized goby, and in having more restricted gill openings and reduced nape scales (see Greenwood (1994) for full descriptions and differences between these species). In the earlier systematic work of the genus, Hoese (1986), also reported another description of a third species *Glossogobius biocellatus* Valenciennes, 1837. This species has since been reassigned to a new genus and is now considered to be *Psammogobius biocellatus* (Valenciennies, 1837) (Bauchot *et al.* 1991). This species (*P. biocellatus*) is restricted and present only in small numbers in estuaries from Port Elizabeth and north-eastwards in South Africa (Hoese 1986). There have therefore been no published systematic studies on these species after the work of Bauchot *et al.* (1991), Hoese (1986) and Hoese and Winterbottom (1979). It is however possible that some species have been inadequately defined such that synonymies like *G. tenuiformis* (see Figure 1.2) could represent a valid species while some widespread species may represent species complexes, because most of the characters currently used to define *Glossogobius* species may not be subjected to evolution and are highly conserved among species. The distinctiveness of many *Glossogobius* species remains difficult and may require a more comprehensive assessment of some characters such as the papillae patterns as well as their genetic makeup.

1.2.1 *Glossogobius callidus*

Glossogobius callidus (river goby) (Figure 1.2), is one of the presently recognised species of *Glossogobius* occurring in southern Africa. It is endemic to southern Africa and is abundant in coastal plain rivers and estuaries. Its distribution ranges from inland Malawi (Tweddle 2007) to the Swartvlei estuary in the Western Cape, South Africa (Figure 1.3) (Whitfield 1998). In South Africa adult specimens of this species have also been recorded from the northern low-veld region in the Limpopo river system (Greenwood 1994). *Glossogobius callidus* exhibits an extensive salinity tolerance as it is found in both estuarine and freshwater habitats (Engelbrecht and Mulder 1998; Whitfield 1998). Members of this species reach sexual maturity at approximately 35 mm standard length (SL) and breed during spring (October - November) in the Eastern Cape region of South Africa (Whitfield 1998). Depending on the biogeographical region, larvae are most abundant during November (late spring), declining gradually over summer (Harris 1996). Larval *G. callidus* in South African waters are found in permanently open and intermittently open estuaries (Strydom and Neira 2006). Juveniles of this species have been collected mainly from the middle and upper reaches of southern African estuaries (Whitfield 1998).

The extensive distribution of this species in both freshwater and estuarine waters has raised the possibility of cryptic species (i.e. species not easily identified using traditional methods). Engelbrecht and Mulder (1998) questioned the species high genetic variation when they examined two *G. callidus* populations from riverine and estuarine systems. However, they were uncertain whether their results indicated two reproductively isolated species within the present *G. callidus* population. Indeed preliminary research including examination of relevant type specimens, of the South African *Glossogobius* species, indicated that *G. callidus*, as reported in the literature is probably a complex of several cryptic species (Helen Larson, *pers. comm.*; Art Gallery and Museum of the Northern Territory). Therefore, a study of the phylogeography of this species in South Africa will contribute to a better understanding of its genetic variation and possible cryptic species composition throughout distribution in our waters.

1.2.2 *Glossogobius giuris*

Glossogobius giuris (Figure 1.2) is a tropical Indo-west Pacific species whose distribution, in southern Africa, ranges from east coast of Africa down to the Eastern Cape province in South Africa (Figure 1.4) (Whitfield 1998; Skelton 2001). In South Africa this species occupies brackish water habitats and floodplain pans and extends hundreds of kilometres inland into the larger rivers (Skelton 1993; Whitfield 1998; Skelton 2001). Although it appears to be mainly a freshwater species, the

largest specimens in southern Africa have been recorded in estuarine lagoons and lakes. Sexual maturity of this species is reached at about 55 mm SL (Whitfield 1998). In contrast to *G. callidus*, breeding occurs during summer. Juveniles in freshwater habitats feed on benthic invertebrates whereas those in estuaries feed on amphipods, decapods, tanaids and chironomid larvae (Van der Elst 1988; Skelton 1993, 2001).

As a consequence of the difficulty in the identification of morphological features in gobies, *G. giuris* for example, may also represent a species complex (see Akihito and Meguro 1975) which may be in need of revision. Even with the wide distribution (from Africa, Australia and Madagascar to India and south of China) of *G. giuris*, only a few morphological differences among the different areas, regions and habitats (i.e. estuarine or freshwater) have been recorded. The major intraspecific differences include the pectoral fin ray counts; cheek papillae row patterns, pit organs and pre-dorsal scale counts (Akihito and Meguri 1975). Thus there is a need to re-evaluate the distinguishing morphological characters for the species or incorporate other taxonomic techniques such as molecular systematics to understand this species. Synonyms of the species such as *G. tenuiformis* (see Figure 1.2) may represent valid southern African species that are in need of re-evaluation.

1.3. Molecular systematics

Molecular systematics can be defined simply as the study of the diversity of organisms and the relationships among them with the use of information from molecular data (Hillis and Moritz 1990). Since the birth of and advances in molecular systematics, interest in morphological data as a phylogenetic tool seems to have declined. However, as stated before, studies using morphology are important in tracing the evolution of any number of specific physical traits and are essential to the field of taxonomy (Harrison 1989; Thacker 2003). A fundamental technical innovation that has made molecular systematics and the use of molecular data a fast growing science was the advent of the polymerase chain reaction (PCR). PCR is a technique that generates hundreds to billions of identical molecules of DNA from a single molecule in just a few hours (Mullis 1990). The size of the data set of informative characters available from molecular data is one of the main advantages of molecular systematics over morphological data (Hillis and Moritz 1990). The large size and diversity in rates of change in different parts of a genome ensure that each of the nucleotide positions in a DNA sequence can be considered a character, which increases the informative data for analyses (Avise 2004). Thus, molecular characters provide a rich source of information for analyses of species phylogeny and, within limits, can indicate the approximate timing of evolutionary events among taxa (Hillis and Moritz 1990).

Studies of molecular systematics have already revealed some astonishing results about the evolutionary interrelationships and phylogeny of many groups of goby taxa (Thacker and Cole 2002; Thacker 2003). Hou *et al*. (2007) suggested that the role of phylogenetics in molecular systematics is to define the ancestor and descendant relationships while providing a hierarchical foundation for species taxonomy. Additionally, Hillis and Moritz (1990) reported that molecular systematics and phylogeny also reveal previously inaccessible details of historical biogeography and population processes that may have caused the present observed species, and phylogeographic structures. Wang *et al*. (2001) used these advantages of molecular systematics to propose the molecular phylogeny of the gobioid fishes using rRNA sequences, while Thacker (2003) used several complete sequences from the mitochondrial DNA. With the same types of taxonomic questions to answer for the gobioid fishes, Akihito *et al*. (2000) hypothesised a new phylogeny at the familial level, while Harold *et al*. (2008) managed to provide a better resolved phylogeny within the genus *Gobiodon* Bleeker, 1856 using molecular data. Thus phylogenetic studies are not only producing insights into taxonomy but also the evolution of adaptation and divergence among biogeographic regions and within species. Molecular studies should therefore provide more characters for distinguishing species of unclear genera like *Glossogobius*

1.4. Genetic structure and phylogeography

Traditionally, population genetics has focused on evolution and the role of evolutionary factors in gene flow and in shaping the genetic structure of populations (Altukhov 2003). Gene flow, defined as the movement of gametes or individuals from one place to another, influences both the population structure and geographic distribution of a species, as well as the adaptation of isolated populations to their local environments (Slatkin and Maddison 1989). The major evolutionary factors that have been identified as influencing the level of differentiation and gene flow between natural populations include, amongst others, dispersal ability, geographic barriers, natural selection and random genetic drift (Avise 2004). These factors affect gene flow and therefore species differentiation at various levels. Hypothetically, the larger the dispersal potential, the greater the gene flow, the less the genetic structure (Avise 2000; Dawson *et al*. 2002).

It has been observed in some instances, that species may show such high levels of gene flow to an extent that they become almost panmictic (i.e. referring to unstructured random-mating populations). For example, this is common in marine and estuarine systems, where it is not easy to identify boundaries as the ocean provides a single continuous dispersal medium that connects all habitats of a

species' distribution. In these two systems (marine and estuarine) most fish species have also been found to have large effective population sizes over extensive geographic ranges, with high levels of gene flow among populations, thus leading to low genetic structuring (von der Heyden *et al*. 2007; Neethling *et al*. 2008). These high levels of genetic similarity among populations observed for a number of fish species are counter to that expected for dispersal by contemporary ocean current systems and may reflect stochastic expansion events (Palumbi 1997). However, some species exhibit very low levels of gene flow suggesting that natural selection and genetic drift may occur almost independently in each local population (Francisco *et al*. 2006). This phenomenon has been mainly observed among fish species that are distributed in freshwater and terrestrial environments, where habitat attributes can change over a short distance as mountains and vegetation create effective barriers that can reduce the dispersal potential of organisms. This isolation in turn may result in more pronounced phylogeographic structure within these species (Swartz *et al*. 2007).

Thus phylogeographic structure, defined briefly as the genetic composition and structure of an extant population over a geographic distribution (Avise 2000) can be used to observe the influence of historical factors on present patterns of biodiversity. By using phylogeography we can also infer the processes that affect the genetic composition of a species, i.e. historical events, such as habitat fragmentation, isolation by distance and/ or range expansion of a species (Avise 2000). Today, a range of DNA techniques, combined with the new analytical methods, and recent paleoclimatic and geological data, are providing more information about the distribution or phylogeography of a species across all of its distribution (Riddle *et al*. 2000; Dawson *et al*. 2001; 2002).

Although most phylogenetic studies explore patterns within single species, greater access to larger amounts of molecular data has made comparable phylogeography (i.e. a study that compares the genetic relationships of two or more species with overlapping geographic distributions) a powerful and increasingly popular subject (Riddle *et al*. 2000; Stefanni and Thorley 2003). Many studies of comparable phylogeography in fishes have revealed common patterns and causes of genetic heterogeneity within species, as well as suggesting possible links between water systems, life histories, and genetic structures (Dawson *et al*. 2001; Swartz *et al*. 2007). Avise *et al*. (1987) proposed three phylogeographic hypotheses (Table 1.2), which have been used in most of the studies available in the literature. More recently, Swartz *et al*. (2007) tested these hypotheses in a comparative study of *Pseudobarbus afer* Peters, 1864 and *Pseudobarbus phlegethon* Barnard, 1938 (Cyprinidae), with the striking result that four divergent lineages existed within *Pseudobarbus afer*. This study (Swartz *et al*. 2007) highlighted the third hypothesis (in Table 1.2) which states that

monophyletic groups distinguished by large phylogenetic gaps usually arise from long-term zoogeographic barriers to gene flow.

1.5. Genetic markers

In the recent years, molecular techniques have been used as a tool for phylogeographic studies by analyzing dispersal, colonisation patterns, and gene flow of populations over a variety of geographic scales (Duran *et al*. 2004). Genealogical relationships of species have also been done to assist in mapping patterns of diversity within and between remnant populations of threatened and endangered species (Vrijenhoek 1998). Following the development of polymerase chain reaction (PCR) technologies and DNA sequencing (Kocher and Stepien 1997; Carvalho and Hauser 1998), the next major impact on systematics and determination of genetic differences came about from the introduction of sensitive molecular markers, most notably the use of mitochondrial (e.g. cytochrome *b* and cytochrome *c* oxidase, subunit 1) and nuclear DNA (e.g. introns and microsatellites) (Fu and Li 1999). These developments of molecular markers and techniques have helped invigorate the systematic studies of fish, particularly of groups that have proved difficult using morphology (Thacker 2003). Because of their evolutionary conserved nature, the above genes are also useful for inferring genetic distance relationships and can provide a means of assessing relationships between organisms that lack any informative homologous morphological characters (Fu and Li 1999; Thacker 2003; Hou *et al*. 2007).

However, patterns of population genetic structure detected using different markers are subjected to various modes of evolution and histories (Wade *et al.* 1994). Detected patterns of population genetic variation among nuclear DNA and mtDNA data sets may differ according to the response of each independent genome to the combined effects of mutations, migration, and effective population size (Hou *et al*. 2007). Thus the rate of approach to genetic equilibrium can vary among these classes of genetic markers due to different mechanisms that are at play in terms of mutations and modes of inheritance. Therefore, it is important to note that different methods of analysis probe different aspects of the molecular and spatial history of an organism (Carvalho and Hauser 1998; Fu and Li 1999). Because each DNA sequence has its own genealogy, the evolutionary history of the organism is therefore the sum of many different gene genealogies. Consequently, one would ideally like to use a range of markers including nuclear, cytoplasmic, conserved and neutral sequences that include examples of high and low mutation rates to reconstruct a species phylogeographic history and phylogenetic relationships (Bowen *et al.* 1993; Harada *et al*. 2002; Roe *et al*. 2002). Systematic

studies using data of genealogies from several independent loci should help to provide a more complete and reliable species history or phylogeny (Posada and Crandall 2001; Hewitt 2004).

Therefore, the trend in systematic studies has been to use different genes in molecular studies i.e. one nuclear, and one mitochondrial to obtain more information about complete pictures of populations and phylogeny (Field *et al*. 1988; Hedges *et al*. 1990; Turbeville *et al.* 1991; Gadagkar *et al*. 2005). These two unrelated regions are commonly selected because they represent both more conservative and highly variable regions, and therefore together provide a good phylogenetic signal for analyses. By exploring these differences, systematic studies can provide a means for testing competing hypotheses of population structure not otherwise possible using single loci (Hey and Harris 1999; Gadagkar *et al*. 2005). This approach has been useful for identifying instances of sex-biased dispersal (Karl *et al*. 1992), testing hypotheses of historical demographic changes (Avise 2004), and estimating effective population sizes (Slade *et al*. 1998).

1.5.1 Animal Mitochondrial DNA marker (mtDNA)

The first DNA-based genetic marker system that could be routinely applied to surveys of genetic variation in natural populations was animal mtDNA (Neigel 1997; Avise 2000). Mitochondrial DNA is a circular molecule, approximately 16-20 kilobases (Kb) in size, with 37 genes encoding 13 proteins, two ribosomal RNAs (rRNA), 22 transfer RNAs (tRNAs), and one major non-coding region (control region) that contains the initial sites for mitochondrial replication and transcription (Figure 1.5) (Brown *et al*. 1982; Meyer 1993; Inoue *et al*. 2000). Mitochondrial DNA sequence analysis has become one of the most widely used tools in studies of molecular phylogeny and phylogeography among vertebrates including fishes because it is easy to handle and can be easily purified and sequenced (Avise 2000). Furthermore, mtDNA is useful in molecular studies due to its pattern of inheritance (maternally inherited, without recombination) (Moritz *et al*. 1987; Meyer 1993; Davies-Coleman *et al*. 2000). Because the mitochondrial genome is generally transmitted without recombination as a single linkage unit, mtDNA sequence variants are usually referred to as haplotypes rather than alleles in population studies (Moritz *et al*. 1987; Meyer 1993). The inheritance of mtDNA is thus formally similar to that of a single haploid locus (Moritz *et al*. 1987).

After more than two decades of using mtDNA in population genetic studies, the advantages and its applications have remained valid for most taxa (Ladoukakis *et al*. 2002). There are, however, some general problems associated with the use of mtDNA (Meyer 1993). Analysis of genetic variation in short sequences have in some cases illustrated ambiguous geographic structures of local populations,

mainly because the sequence amplified was either too short to contain variation, or the evolutionary rate of the segment was not suitable for the specific purpose of the study (Inoue *et al*. 2000; Mindel *et al*. 1999). Heteroplasy (when two or more genotypes coexist within the same individual) is another limiting factor of using mtDNA as a genetic marker. However, this has little impact, since such cases are unusual and rare (Avise *et al*. 1987; Meyer 1993). Thus, an accurate investigation of spatial or temporal genetic structure or phylogenetics should consider the variety of patterns seen in different loci as phylogenetic trees derived from a single marker may not accurately reflect the genetic history of a population, species or genus (Slatkin and Maddison 1989; Palumbi and Baker 1994; Gadagkar *et al*. 2005). It is important to note that different gene fragments of the mtDNA region have different evolutionary rates and have been used to answer different phylogenetic questions.

1.5.1.1 Cytochrome *b*

The cytochrome *b* region of mtDNA (Figure 1.5) is one of the most extensively sequenced and widely used mitochondrial genes for phylogenetic work on fishes due to several reasons. The evolutionary dynamics and the biochemistry of the protein product are better characterised than most other molecular systems. Cytochrome *b* is also one of the cytochromes involved in electron transport in the respiratory chain of the mitochondria and is the only cytochrome coded by mtDNA (Irwin *et al*. 1991). Finally, levels of genetic divergence typically associated with sister species are usually phylogenetically informative in cytochrome *b* analyses and unlikely to be severely compromised by saturation effects involving superimposed nucleotide substitutions (Unseld *et al*. 1995). In fishes, the cytochrome *b* region is known to be variable enough for use in studies at the population level (Birt-Friesten *et al*. 1992; Dawson *et al*. 2001; 2002; Roe *et al*. 2002), and conserved well enough for clarifying deeper phylogenies at generic and family levels (Meyer *et al*. 1990).

Phylogenetic studies of fishes using cytochrome *b* are very common for chondrichthyan fishes (Martin and Palumbi 1993), teleost fishes (Bartlett and Davidson 1991; Block *et al*. 1993) as well as the larger perciform group of fishes (Cantatore *et al*. 1994). The cytochrome *b* gene is under strong evolutionary constraints with some parts of the gene being more conserved than others due to functional restrictions (Meyer 1994). Although uses of cytochrome *b* have some pitfalls, authors, such as Meyer (1994), argued that it is also the best choice for resolving relatively recent evolutionary histories. Harada *et al*. (2002) used the cytochrome *b* gene in estimating the phylogenetic relationships of four species of floating gobies (*Gymnogobius* Gill, 1863) in Japan and Korea while Akihito *et al*. (2000) also established the phylogenetic relationships of the gobioid fishes, at the familial level. However, it is generally agreed that a nuclear marker should also be used

to determine paternal gene-flow patterns since cytochrome *b*'s maternal mode of inheritance allows only the inferences of maternal patterns of gene flow (Gadagkar *et al*. 2005).

1.5.2 Nuclear DNA marker

The nuclear genome provides an incredible diversity of markers available to evolutionary scientists. The genome has both coding and non-coding regions that evolve at different rates, allowing a broad range of inferences from intra-population dynamics based on markers from independent linkage groups (Sunnucks 2000; Fujita *et al*. 2004). The nuclear DNA polymorphisms that exist widely in eukaryotic organisms provide virtually unlimited opportunities for studying the mechanisms of evolution. Although it is widely known that the rate of evolution of nuclear sequences is much lower than that of mtDNA in many vertebrates (Brown *et al*. 1982), it has been reported that this is not a general rule (Zhang and Hewitt 2003). A large proportion of nuclear DNA markers employed for population analyses have been the non-coding regions, because they tend to be more variable than the coding regions (Zhang and Hewitt 2003). Apart from nucleotide substitutions, insertions and deletions (indels) often constitute a large part of the polymorphism detected (Young and Healy 2003). Thus the use of data contained in gaps is important for the interpretation and exploitation of polymorphism data from nuclear markers as the pattern of indels contains a sufficiently large amount of phylogenetic signal (Simmons and Ochetera 2000; Zhang and Hewitt 2003).

There are relatively few studies (Schluter 2000; Steiner *et al*. 2005) that have estimated gene flow using nuclear DNA (nDNA), employing either variable number of tandem repeats (VNTR's, e.g. ribosomal RNA genes (rDNA), microsatellites) and/or base substitutions (e.g. introns) (Neigel 1997). Analyses of nuclear markers, such as introns and microsatellites provide both paternal and maternal information (Neigel 1997). Nuclear introns have been used as additional sources of genetic variation for species-level phylogenies since they are abundant throughout the genome (Steiner *et al*. 2005). It has been observed that nuclear introns tend to have an elevated rate of evolution when compared to coding sequences (Palumbi and Baker 1994; Gaffney 2000; Armstrong *et al*. 2001). Several previous studies on fish have applied nuclear DNA in phylogeny estimation (e.g. Carlsson *et al*. 2004; Lavoue *et al*. 2003; Palumbi and Baker 1994). However, it is very important to consider recombination, selection (non-neutrality), heterozygosity, insertions/deletions polymorphism (indels), low divergence and gene-specific variation in rate and evolution history when working with nuclear markers (Zhang and Hewitt 2003).

2. AIMS AND OBJECTIVES

The overall aim of the current study was to examine the systematics of *Glossogobius* in southern Africa using molecular markers (mtDNA cytochrome *b* and nuclear S7 intron 1) in an attempt to clarify the biogeographic distribution and species status.

The two specific objectives of the study are:

1. To estimate of the phylogeographic structure of *G. callidus* using both mtDNA (cytochrome *b*) and nuclear genes (S7 gene introns 1) (Chapter three).

The wide distribution covering different habitats and biogeographical zones as well as the high levels of morphological variation observed within *G. callidus* make it a good study species from a phylogeographic perspective. Knowledge of the phylogeography and genetic variation within the species will not only help solve the dilemma of cryptic species but also provide some information on the possible historical geographic and genetic connections between the localised populations within the river and estuarine systems. Even though the study of Engelbrecht and Mulder (1998) suggests that genetic variation should be expected, a low genetic structuring can also be expected if there are high levels of dispersal among its estuarine populations. Thus, the objective of this study was to do a large-scale assessment of population structure, genetic diversity and phylogeography of *Glossogobius callidus* in southern Africa. Studying the genetic variation within this widespread species will help us investigate the distribution and drainage history as well as the genetic variation of the species within the two water habitats throughout its distribution.

2. To examine the species relationships and the evolutionary history of the flat head gobies found in southern African waters, using the cytochrome *b* and nuclear S7 intron 1 genetic markers (chapter four).

As previously discussed, the lack of distinguishing morphological characters within the genus has resulted in a number of previous species descriptions being synonymised. Although *Glossogobius tenuiformis* was synonymised with *G. giuris*, this synonym could be a valid African *Glossogobius* species. Considering the wide distribution range from Africa to Australia as well as the extent of observed morphological characters within *G. giuris*, this species may also represent a species complex. The other southern African species *G. callidus* is also very variable morphologically and

genetically and is considered to be a species complex. There is therefore a need to investigate and re-evaluate the species' relationships of the flat head gobies found in South African waters.

This study thus aims to identify the phylogeographic structure and processes (river capture, sea level changes, and migration within systems) that have played an important role in shaping the genetic patterns within *G. callidus*, and to hypothesise the relationships of the genus *Glossogobius* in southern Africa. It will also be necessary to verify whether genetic data can be useful in evaluating and understanding diversity within this genus.

Table 1.1. Valid species of the genus *Glossogobius* (N = 22) according to FISHBASE (www.fishbase.org).

No.	Valid Name	Author	Common Name
1.	*G. ankaranensis*	Banister, 1994	
2.	*G. aureus*	Akihito and Meguro, 1975	Golden tank goby
3.	*G. bicirrhosus*	Weber, 1894	
4.	*G. brunnoides*	Nichols, 1951	Dusky mountain goby
5.	*G. bulmeri*	Whitley, 1959	Bulmer's goby
6.	***G. callidus***	Smith, 1937	River goby
7.	*G. celebius*	Valenciennes, 1837	Celebes goby
8.	*G. circumspectus*	Macleay, 1883	Circumspect goby
9.	*G. coatesi*	Hoese and Allen, 1990	Coates' goby
10.	*G. concavifrons*	Ramsay and Ogilby, 1886	Concave goby
11.	*G. flavipinnis*	Aurich, 1938	
12.	***G. giuris***	Hamilton, 1822	Tank goby
13.	*G. hoesei*	Allen and Boeseman, 1982	Hoese's goby
14.	*G. intermedius*	Aurich, 1938	
15.	*G. kokius*	Valenciennes, 1837	
16.	*G. koragensis*	Herre, 1935	Koragu tank goby
17.	*G. mas*	Hora, 1923	
18.	*G. matanensis*	Weber, 1913	
19.	*G. obscuripinnis*	Peters, 1868	
20.	*G. olivaceus*	Temminck and Schlegel, 1845	
21.	*G. sparsipapillus*	Akihito and Meguro, 1976	Linecheek tank goby
22.	*G. torrentis*	Hoese and Allen, 1990	White water goby

Table 1.2. Intra-specific phylogeographic hypotheses (Avise *et al*. 1987).

Hypothesis 1: Most species are composed of geographical populations whose members occupy different branches of an intraspecific, phylogenetic tree.

Hypothesis 2: Species with limited phylogeographic population structure have life histories conductive to dispersal and have occupied ranges free of firm impediments to gene flow.

Hypothesis 3: Monophyletic groups distinguished by large phylogenetic gaps usually arise from long-term extrinsic barriers to gene flow.

- As time since isolation increases, the degree of phylogeographic concordance across separate genealogies increases.
- The geographical placements of phylogenetic gaps are concordant across species.
- Phylogenetic gaps within species are geographically concordant with boundaries between traditionally recognised zoogeographic boundaries.

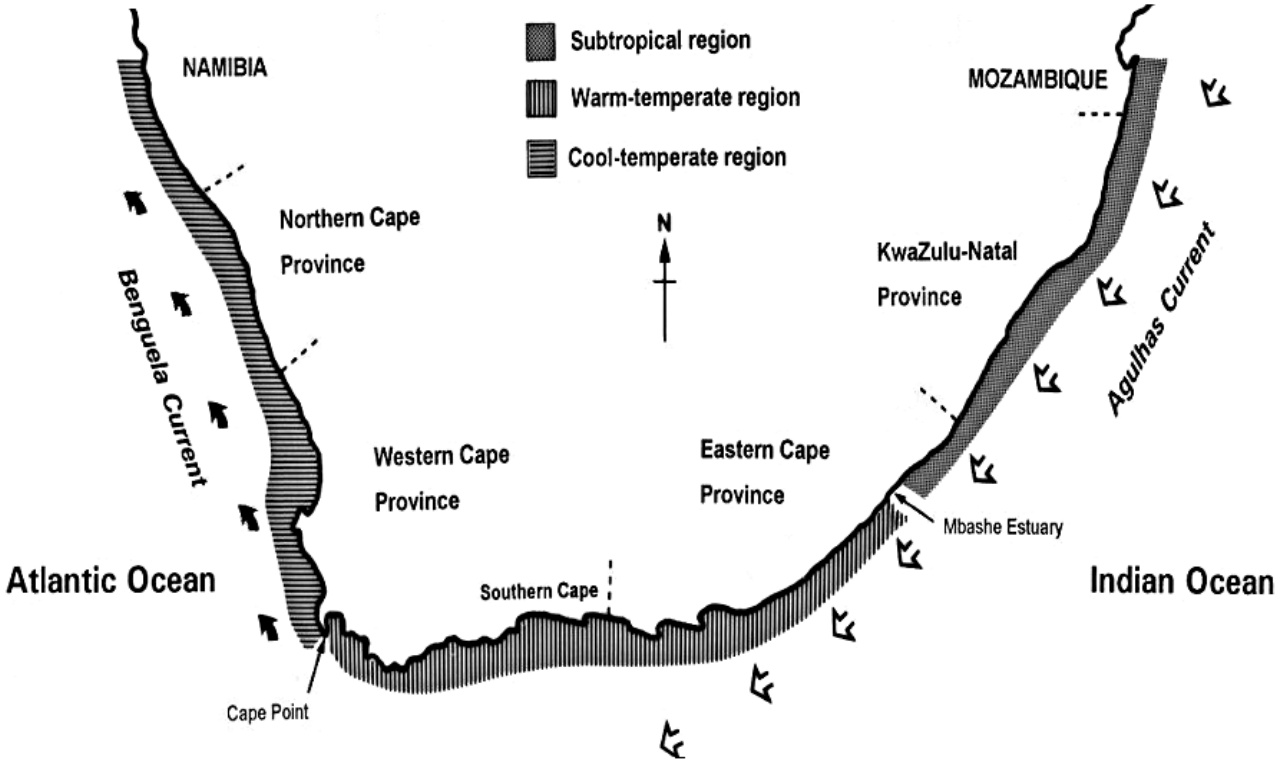


Figure 1.1. Map of South Africa indicating three biogeographic provinces along the coastline (after Whitfield 2000). Arrows at Cape Point and Mbashe Estuary indicate the breaks between the different biogeographic zones. Also shown is the main current system (warm Agulhas) that influences both the warm-temperate and the subtropical regions on the South African coastline.

Figure 1.2. Colour photographs of *Glossogobius callidus* (Smith, 1937) (top, Ofer Gon (SAIAB), *Glossogobius giuris* Hamilton, 1822 (middle, www.siamensis.org/.../ FishesPics_reply_24329.jpg), and *Glossogobius tenuiformis* Fowler, 1934 (bottom, Ofer Gon, SAIAB).

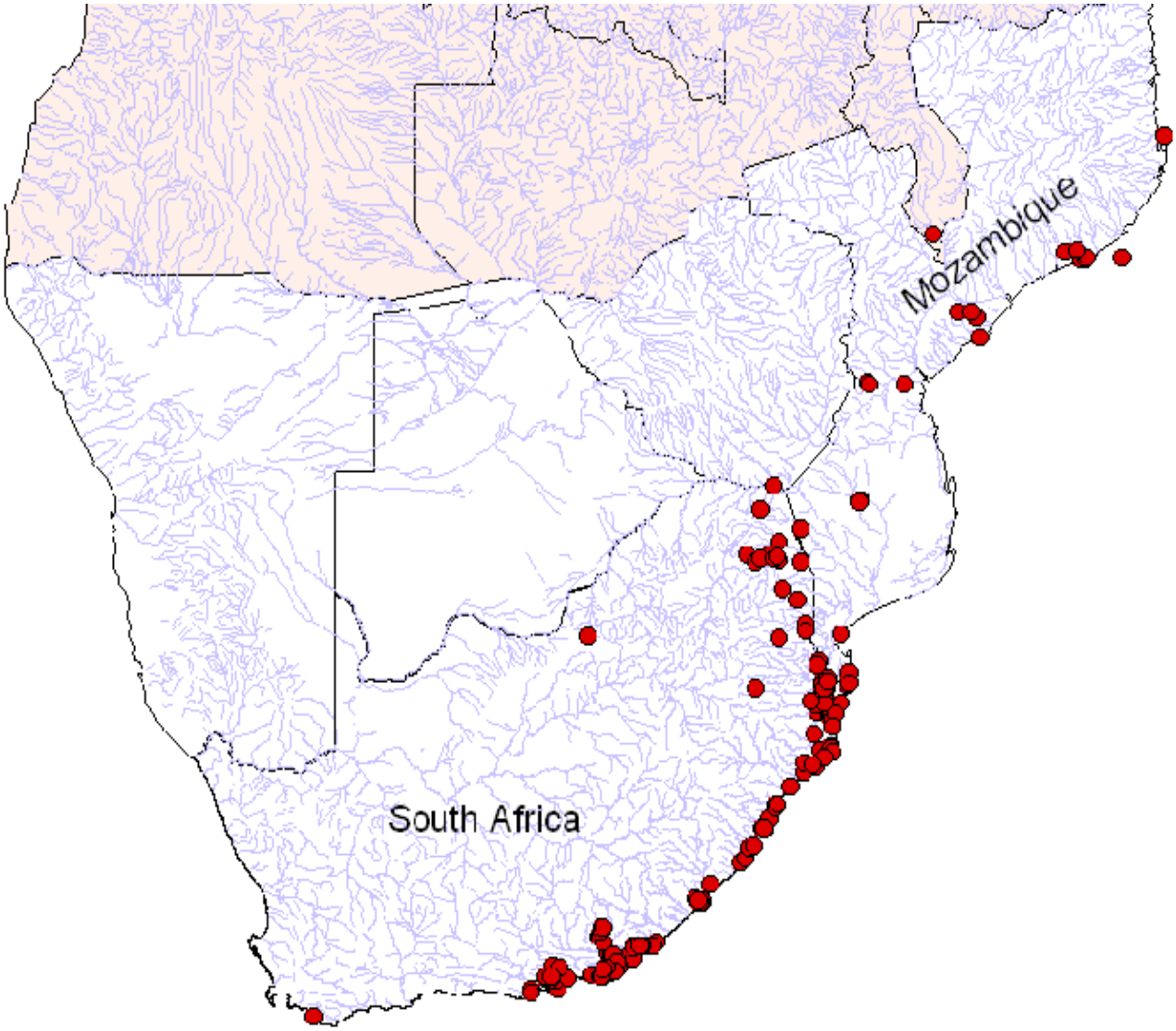


Figure 1.3. Distributional range of *Glossogobius callidus* (Smith 1937) in southern Africa illustrating estuarine and river occurrences. Distribution starts from Mozambique to the Eastern Cape Province in South Africa. Red circles show the recorded localities. This information is provided from the SAIAB national fish collection records.

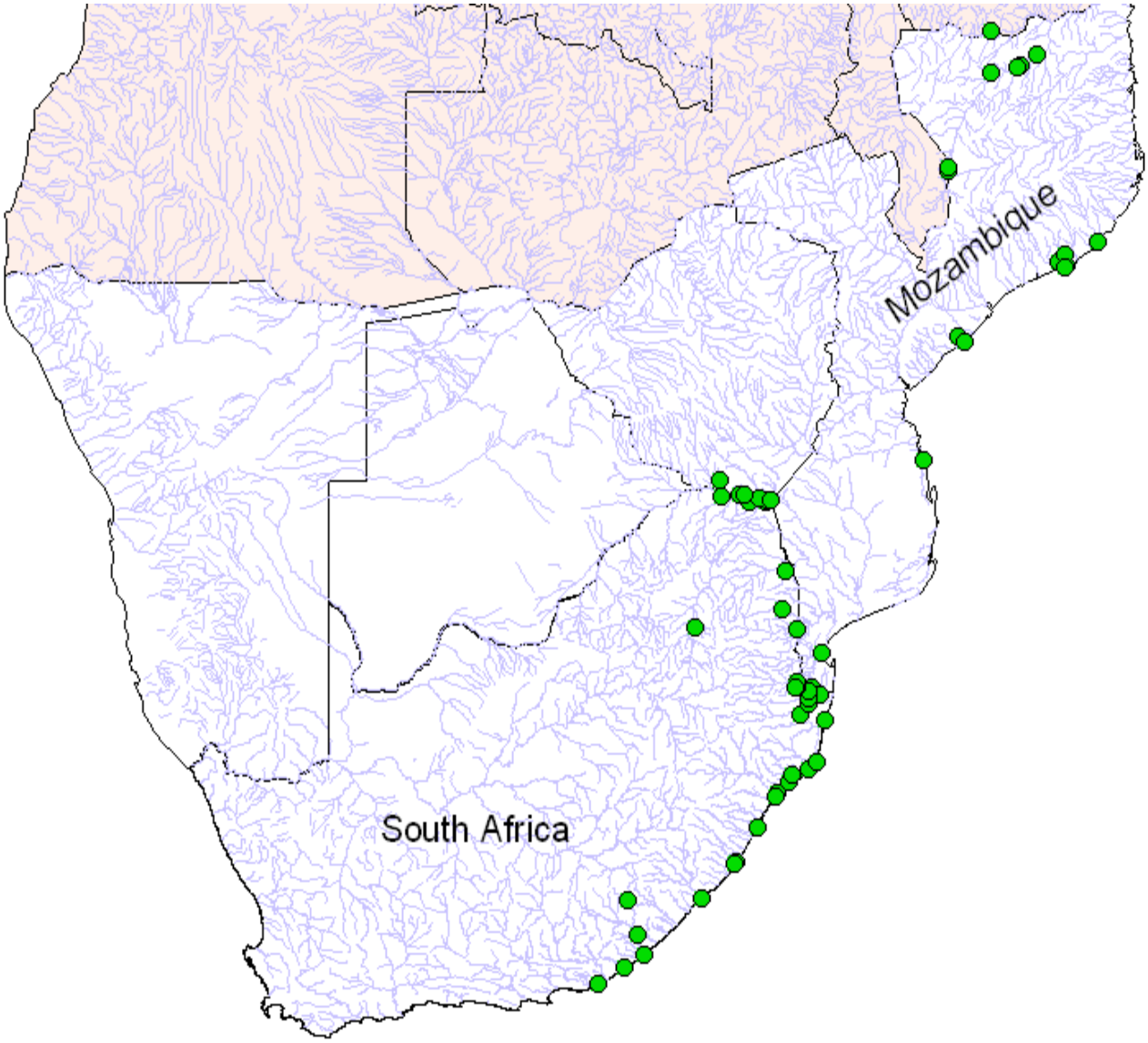


Figure 1.4. Southern African distributional range of *Glossogobius giuris* Hamilton, 1822 illustrating estuarine and major river occurrences in the southern Africa region. Distribution starts from the upper Mozambique down to the Eastern Cape Province in South Africa. Green circles show recorded localities. This information is provided from the SAIAB national fish collection records.

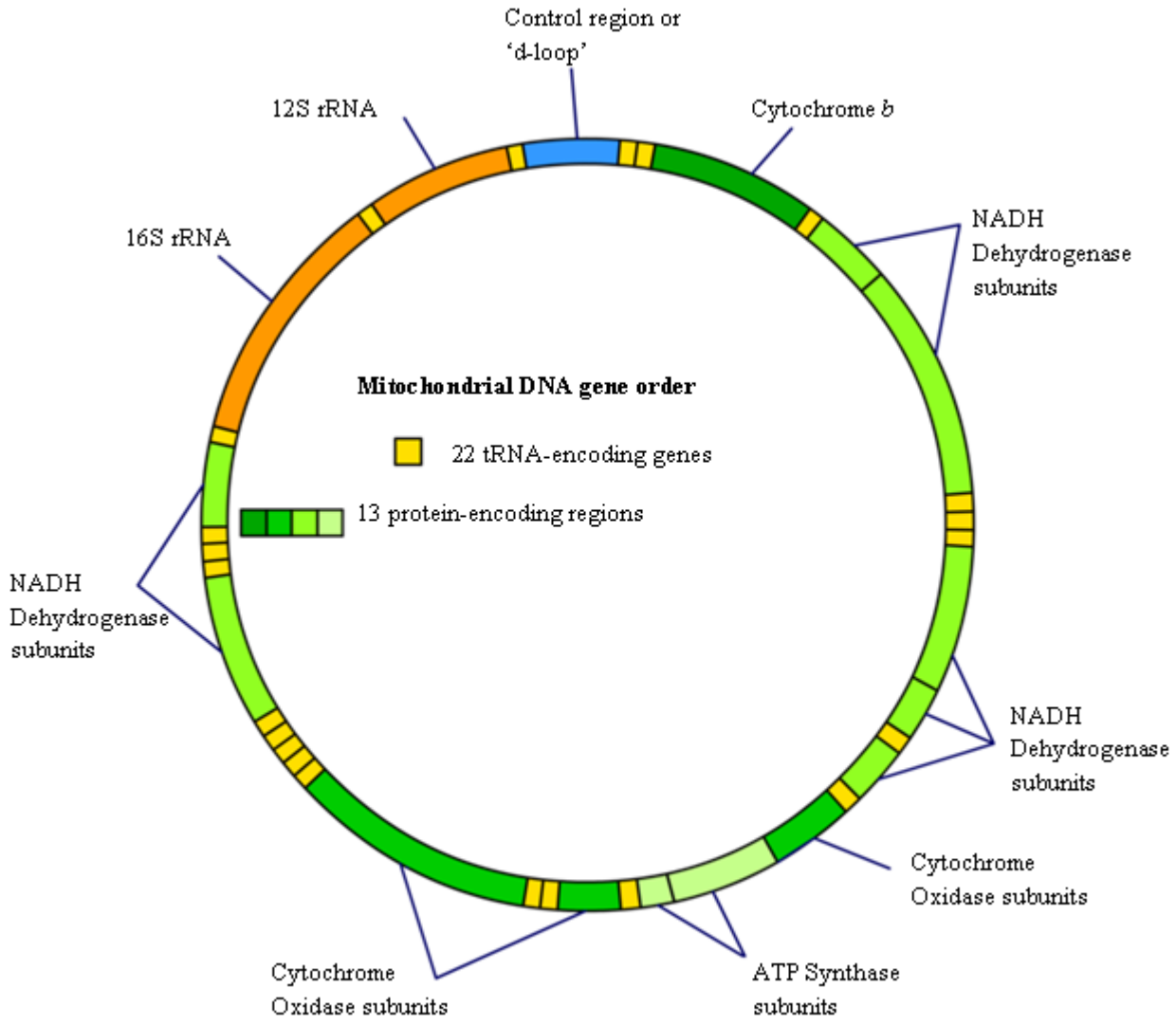


Figure 1.5. Animal mitochondrial gene order, illustrating the different regions, including the cytochrome *b* region in fish that was used in the study (after Meyer 1993; Source= http://commons.wikimedia.org/wiki/Image:Mitochondrial_DNA_blank.png).

CHAPTER 2

MATERIALS AND METHODS

2. OVERVIEW

This chapter describes the materials and methods used to study the phylogeography of *Glossogobius callidus* (Chapter 3) and the systematics of southern African species of *Glossogobius* (Chapter 4). Samples used for both studies were obtained from locations across the species' distribution in southern Africa (Table 2.1), and with the addition of a few samples from Australia donated by the Art Gallery and Museum of the Northern Territory (NTM), Darwin (Table 2.2). Methods used included direct sequencing of mitochondrial DNA (mtDNA), cytochrome *b* (cyt *b*) and nuclear genes (nuclear S7 ribosomal protein gene intron 1), relevant phylogenetic statistical procedures and morphological assessment using both light and scanning electron microscopy (SEM)

2.1. Sampling procedure and species identification

In South Africa, individuals were collected from rivers and estuaries throughout the species' distribution in both biogeographic regions (subtropical and warm-temperate) using 6 meter seine nets (Figure 2.1 and Table 2.1). In the field, specimens were frozen or fixed in 95% ethanol and labelled with station numbers. Additional specimens of both *G. callidus* and *G. giuris,* collected from Mozambique and Swaziland and included in the study, were already available from the SAIAB fish collection facility (Table 2.1). Australian specimens of the widespread *G. giuris, G. aureus* (Asia and Oceania: Japan to Australia) and unidentified *Glossogobius* species from Australia were also used for the phylogenetic analyses. Outgroup sequences were downloaded from Genbank (Table 2.2). These sequences were selected from available goby sequences on Genbank that showed the highest percentage similarity to the sequenced *Glossogobius* sequences in this study.

Using the key from the work of Helen Larson and associates (*pers. comm.*), collected individual fish were identified to the species level in the laboratory. *Glossogobius* specimens that fitted the description of *Glossogobius tenuiformis* were also identified (and then treated as *G. tenuiformis* in subsequent analyses). All the identifications were done in the laboratory at the South African Institute for Aquatic Biodiversity (SAIAB) using a dissecting microscope. Further morphological work was done with the aid of a Scanning Electron Microscope (SEM) at the Electron Microscopy Unit of Rhodes University, Grahamstown (South Africa). Prior to the scanning, the right cheek of the

fish was carefully removed to minimize chances of damaging any pappilae structures. The specimens were prepared for sputter-coating (gold-palladium) by dehydrating them through a graded ethanol series (70, 80, 90, 100% immersed in each for 30 min), followed by critical point drying (CPD), where after they were mounted. Twelve specimens were used for this procedure.These specimens were randomly chosen from the collected localities which represented both biogeographic provinces in the study area. For *G. callidus* chosen localities were Bushman's (type locality), Kariega and Gonubie from warm-temperate region, and Mzingazi canal, Mzimayi, Mhlatuzi and Shengeze Lake from the subtropical region. Only two localities (Mzimvubu and Matigulu, collected during the course of the study) and a specimen number SAIAB 36001 from the SAIAB fish collection were selected for *G. tenuiformis*. Two more specimens representing *G. giuris* (accession numbers: SAIAB30517 and SAIAB73983) where also taken from the fish collection.

2.2. DNA extractions

Muscle tissue from individuals was subjected to total DNA extraction following the protocol of the Promega DNA purification kit (Madison, New York). This process involved an initial incubation of the samples in an extraction buffer with added proteinase K at 55 ^{o}C and intermittent vortexing for 3–4 hours or overnight. Extracted DNA was visualised by loading 5 µl of DNA template mixed with 2 µl of Bromophenol blue on a 1% agarose gel containing Ethidium Bromide. Gels were electrophoresed for 15–20 minutes at 80 mV while submerged in 1X TBE buffer. Each gel was visualised under ultra-violet light (UVP Translluminator), to verify the quality of DNA extracted. DNA extraction aliquots were labelled and stored in a -20 ^{o}C freezer.

2.3. Polymerase chain reaction (PCR)

For the amplification of both genes (cytochrome *b* 1064 base pairs and S7 intron1 857 base pairs), each PCR reaction was made up to the total volume of 50 µl in a 200 µl PCR tube. Each reaction mixture contained: 2–5 µl template DNA; 10X reaction buffer (5 µl), 25 mM $MgCl_2$ (5 µl); 0.125 mM of dNTP's (Promega) (5 µl); 1 units of Super-therm® Taq DNA polymerase (0.2 µl); 25 or 50 *p*mol (1 µl) of the forward and reverse primers of the regions being amplified (Table. 2.2); and autoclaved DNAase free water to top up the mixture to 50 µl. Negative controls without template were included in all PCR reactions. PCR amplifications were performed in either a Thermo Hybaid PX2 or Multi-Block Thermal cycler machine (Southern Cross Biotechnology) under the following conditions:

1) Cytochrome *b*: Initial denaturation at 94 ºC for 3 minutes, followed by 35 cycles of amplification made up of 30 seconds of denaturation at 94 ºC, 30 seconds of annealing at 56.5 ºC, and 90 seconds of extension at 72 ºC. The reaction was completed with a final extension stage at 72 ºC for 10 minutes (1 cycle).

2) S7 intron 1: Initial denaturation at 95 ºC for 8 minutes, followed by 35 cycles of amplification made up of 60 seconds of denaturation at 95 ºC, 60 seconds of annealing at 63-65 ºC, and 90 seconds of extension at 72 ºC. The reaction was completed with a final extension stage at 72 ºC for 7 minutes (1 cycle).

All polymerase chain reaction products were visualised on 1% agarose gels to verify whether the target region had been amplified and to check for possible contamination by checking the negative control.

2.4. DNA sequencing

Prior to sequencing, PCR products were purified using a "QIAquick PCR purification kit" (Qiagen) following the manufacturer's instructions. After a successful PCR product purification, the clean products were then eluted from the cleaning tube in a final volume of 30 µl in a 1.5 ml microcentrifuge tube using 0.1 M TE buffer, pH 8.0. PCR cycle sequencing was conducted using the products and protocols of ABI Prism® BigDye Terminator cycle sequencing ready reaction kit v3.1 (Applied Biosystems). The reaction mixture contained 2 µl purified DNA, 1 µl of either the forward or reverse primer (Note: the primers used for the sequencing were the same as those used for PCR, but were diluted to 5 *p*mol), 3 µl buffer, 2 µl BigDye Terminator cycle sequencing ready reaction (quarter reaction), and autoclaved DNAase free water (12 µl) to make up the final volume to of 20 µl. The reactions were cycle sequenced in a thermocycler using the following conditions: 25 cycles of 10 sec at 96 ºC, 5 sec at 50 ºC and 4 min at 60 ºC.

After cycle sequencing, the tubes were removed from the thermocycler and placed at 4 ºC (to avoid evaporation of the sample) until ready for precipitation. The products were precipitated using Ammonium Sodium Acetate/ EtOH precipitation method (EDTA/ NaOAc/ EtOH) (Applied Biosystems). For each sequencing reaction a 1.5 ml microcentrifuge tube containing the following: 1 µl of 0.25 M EDTA (pH8), 1 µl 3 M NaOAc (pH 4.6), and 50 µl 99 % EtOH were mixed and added to each cycle sequencing product. The mixture was then briefly vortexed and left for precipitation at room temperature for 30 minutes before being centrifuged at 13000 rpm for an additional 20

minutes. The supernatant was removed and discarded while the resultant DNA pellets were washed with 250 µl 70 % EtOH and spun down for 15 minutes at 13000 rpm. After the supernatant was discarded, the resulting pellet was air dried or on a heating block at 60 ºC for 5 - 10 minutes. The properly dried cycle sequencing products were finally analysed on an ABI Prism 310 or 3100 DNA Genetic Sequencers (Applied Biosystems) operated by the Science Faculty, Rhodes University and sequencing facilities of Macrogen (South Korea).

2.5. Statistical analyses

2.5.1 Sequence alignment

Prior to all analyses, all sequences (forward and reverse) were checked and edited for nucleotide ambiguities by eye in a multiple sequence editor, Lasergene v.7.2.0 sequence navigator (DNA Star, Inc., Madison, WI). Multiple sequence alignment of consensus sequences (created in BioEdit software package) were then performed using ClustalX version 1.1 (Thompson *et al*. 1997).

Insertions and deletions (called indels) are common events in the evolution of non-protein coding nuclear DNA sequences such as the nuclear S7 ribosomal protein gene intron 1 used in this study. However, indel events represent two main challenges in phylogenetic analysis of DNA sequences data: position homology (i.e. largely dependant on alignment) and indel treatment as characters (Giribet and Wheeler 1999). This is because indels are not observable characters like nucleotide bases but gaps inserted during sequence alignments to accommodate homologous DNA sequences of unequal length, defining the putative homologous characters amenable to phylogenetic analysis (Simmons and Ochoterena 2000). Because of these problems, a common approach is to avoid or discard regions that have experienced these events. However, gaps can contain important phylogenetic information on the evolution of a lineage that can have dramatic effects on tree topology and clade support (Simmons *et al*. 2001).

Gaps have generally been incorporated into analyses by considering them as either missing data or as another character state in addition to the four nucleotides (i.e., gaps as fifth state). Simmons and Ocheterena (2000) have strongly criticised treating gaps as fifth state because gaps constitute no alternative form of bases but an essentially different form of change or character evolution. Furthermore scoring gaps as fifth state can also result in treating contiguous gap positions as multiple independent characters although they are in most cases considered parsimoniously as a single indel event (Simmons *et al*. 2001). Simmons and Ocheterena (2000) and Simmons *et al*. 2001

recommended that indels be scored as additional absence/present characters, according to the set of rules based on gap overlaps and sharing of the 5' and/or the 3' termini. Therefore, indels were incorporated into the phylogenetic tree analyses by considering them as an additional absent/present character. All indels were coded in SeqState 1.4 (Müller 2005) using the modified complex coding procedure as implemented by Simmons and Ocheterena (2000).

2.5.2 Nucleotide and sequence diversity estimates

To address current and recent historical levels of variation, genetic diversity and effective population sizes within *Glossogobius callidus*, several methods were used as implemented in DnaSP version 4.10.3 (Rozas *et al*. 2003). Sequences were collapsed to unique haplotypes, and genetic diversity was assessed by calculating nucleotide (π) and haplotype (h) diversity within the whole distribution area, and within each locality. Nucleotide diversity was determined as the average weighted sequence divergence between the haplotypes (Nei and Li 1979). Haplotype diversity varies from zero to one, as a measure of the frequency and numbers of haplotypes among individuals (Nei and Tajima 1980) and is the probability of two randomly drawn sequences being different in state. After coding indels as either present or absent in the S7 nuclear intron dataset, the indels were removed from the data to collapse sequences to haplotypes. DNA divergences within species were also calculated using DnaSP version 4.10.3.

2.5.3 Mismatch distribution and neutrality

Observed frequencies of pairwise nucleotide differences (mismatch distribution) were calculated for both gene sequences (cytochrome *b* and S7 intron 1) to test their distribution against a Poisson distribution (Harpending 1994) using the DnaSP package. Mismatch distributions that approximate a Poisson distribution can suggest rapid population expansion (Harpending 1994); whereas stable populations are expected to show a ragged or erratic mismatch distribution. A raggedness index (r) (Harpending 1994) was calculated in DnaSP to quantify the "smoothness" of mismatch distributions. A lower value of r is indicative of a "smooth" mismatch distribution, which in turn could suggest population expansion.

Within DnaSP, a sum of square deviation test was used to test for a fit to a model of sudden population expansion. Using the infinite sites model, the tests were tested for significance using parametric bootstrapping (Schneider and Excoffier 1999). Tajima's *D* (Tajima 1989) and Fu's Fs (Fu 1997) tests were performed and also tested for significance using 1000 bootstrap replicates. Apart from confirming the neutrality of the gene, significant negative values for these tests may also be

interpreted as indicative of sudden population expansions (Tajima 1989). Tajima's *D* value can also be interpreted together with mismatch distribution results, since a negative *D* value predicts a model of sudden expansion while positive values of *D* indicate possible balancing selection of a population (Tajima 1989). Negative *D* values could also suggest recent directional selection, population bottlenecks, or purifying selection on deleterious alleles. Fu (1997) assumes that a negative *Fs* value depicted by a population indicates neutrality of mutations, which are a result of deviations caused by population growth and/or selection. Alternatively, positive values of these test statistics tend to be produced with population subdivisions, an effect that becomes greater the more ancient the division being examined is (Simonsen *et al.* 1995).

2.5.4 Phylogenetic analyses

Maximum likelihood (ML) and parsimony (MP) were performed in PAUP* version 4.0b10 (Swofford 1999) to reconstruct relationships among historically isolated lineages of *Glossogobius callidus* (Chapter 3), and relationships between species (Chapter 4). In addition, Bayesian analyses were performed in MrBayes 3.1.1 (Huelsenbeck and Ronquist 2001) to verify the phylogenetic relationships among southern African *Glossogobius* species (Chapter 4). The model of evolution that best described the data sets, as determined by Akaike Information Criterion (AIC) within Modeltest version 3.06 (Posada and Crandal 1998), was incorporated into the reconstruction of the ML tree. In the phylogeography of *G. callidus* (Chapter 3), two related *Glossogobius* species (*G. tenuiformis* and *G. giuris*) from southern Africa were used as outgroups. Only unique haplotypes were used for all tree constructions in Chapter 3 to reduce computational time. *Psammogobius* species, one from Australia and others from South Africa, and other goby genera found in Genbank (Table 2.2) were used as outgroups (see section 2.1) in the assessment of the phylogenetic relationships of the flat head gobies (Chapter 4). Individuals of all species represented by both genes were used for these analyses. All phylogenetic analyses were performed on cytochrome *b* and S7 intron 1 separately and there after combined, using PAUP and MrBayes software packages.

MP and ML analyses used heuristic searches with 100 and 10 random additional replicates respectively using the tree bisection-reconstruction (TBR) algorithm. Bootstrap values (Felsenstein 1985) for MP tree constructions were calculated using a full heuristic search, simple stepwise-additional option, TBR, and 1000 bootstrap replicates. The model of evolution determined by Modeltest was defined for the ML analyses. In Bayesian analyses, four Markov Chain Monte Carlo (MCMC) chains (one cold and three heated) beginning with random starting trees, and sampling every 500 generations up to 1 million generations. To ensure that all the burn-in trees were removed,

the first 500 trees were discarded from the analyses. A 50% majority rule consensus tree was generated from the trees retained from each run, with posterior probabilities for each node estimated by the percentage of times that the node was recovered.

2.5.5 Population structure and Analysis of Molecular Variance (AMOVA)

In order to assess the distribution of genetic variation within and among collection sites and regions, an analysis of molecular variance (AMOVA) was performed on the identified haplotypes to generate pairwise Φ_{ST}, an analogue to Wright's *F*-statistic (F_{ST}) using Arlequin 3.1 (Excoffier *et al*. 1992; Schneider *et al*. 2000). These *F* statistics were tested for significance through permutation tests (10 000 replicates). An AMOVA can test whether there are different levels of hierarchical genetic subdivisions within a species, i.e. at the level of the individual (within populations in our case localities), between populations and between groups of populations. Φ-statistics hence are the reflection of the correlation of haplotype diversity at these different levels (Excoffier *et al.* 1992). Φ_{ST} values are calculated and defined as the pairwise genetic variation between localities, while Φ_{CT} is defined as the proportion of variation between groups of localities within the total sample. The pairwise Φ_{ST} value does not only take haplotype frequencies into account to estimate variance, but also considers the nucleotide differences between haplotypes. The Tamura-Nei model of substitution with the gamma correction estimated by Modeltest 3.06 was used to calculate genetic distances on which Φ_{ST} calculations was based. Applying different models or gamma corrections did not seem to affect the Φ_{ST} estimates. The last Φ- statistic (Φ_{SC}) estimate is defined as the proportion of genetic variation between localities within groups (Excoffier *et al.* 1992). For the cytochrome *b* data the three hierarchical group structures defined were: (1) Nkanini and Kosi Bay group; (2) Mzimayi group, and (3) the type locality Bushman's group that included the rest of the locations (see chapter 3).

A cladistic analysis of the geographic distribution of *Glossogobius callidus* haplotypes was performed to separate population structure from population history (Chapter 3). Median joining network analysis can discriminate between phylogeographic associations due to ongoing restricted gene flow and those due to historical range expansion, fragmentation or colonisation (Templeton *et al*. 1995). Haplotype networks were constructed for only the cytochrome *b* data sets. The geographic patterns in the relationships among haplotypes and their frequencies in each sampling area were performed using parsimony (MP) in Network v.5.0 package (Polzin and Daneschmand 2003). In this program the outgroup probabilities are assigned to each haplotype based on the correlation between haplotype frequency and the age of the haplotype in the lineage (Templeton *et al*. 1995).

Table 2.1. Localities, SAIAB (South African Institute for Aquatic Biodiversity) numbers and sample sizes for all the collected southern African species of *Glossogobius* used in this study. Warm-temperate (WT), Subtropical (ST) and Tropical (T) biogeographic regions are indicated for the South African species and populations. Eastern Cape (EC) and KwaZulu-Natal (KZN) biogeographic region are indicated.

Locality names	Region	Sample size	GPS positions	Biogeographic region	SAIAB number
***G. callidus*, S. Africa**					
Sundays river	EC	9	33 22.31'S; 25 40.45'E	WT	
Bushman's river mouth	EC	10	33 38.46'S; 26 33.59'E	WT	
Kariega river	EC	16	33 36.55'S; 26 39.27'E	WT	
Kowie estuary	EC	12	33 34.88'S; 26 51.76'E	WT	
West Kleinemond estuary	EC	9	33 32.28'S; 27 02'51E	WT	
East Kleinemond estuary	EC	8	33 32.00'S; 27 02.00'E	WT	
Great Fish river	EC	10	33 10.18'S; 26 54.53'E	WT	
Keisekamma estuary	EC	15	33 11.09'S; 27 23.56'E	WT	
Gonubie estuary	EC	6	32 55.59'S; 28 01.59'E	WT	
Mzimvubu river	EC	4	31 35.50'S; 29 29.59'E	ST	
Mpenjati river	KZN	7	30 59.50'S; 30 15.62'E	ST	
Zotcha estuary	KZN	5	30 46.66'S; 30 22.67'E	ST	
Mtentweni river	KZN	11	30 40.12'S; 30 24.96'E	ST	
Mzimayi river	KZN	7	30 44.21'S; 30 39.54'E	ST	
Mdloti estuary	KZN	5	29 38.78'S; 31 09.30'E	ST	
Mhlatuzi estuary	KZN	10	28 46.86'S; 31 58.59'E	ST	
Mzingazi canal	KZN	2	28 46.13'S; 32 05.61'E	ST	
Lake Mzingazi	KZN	6	28 39.81'S; 32 04.90'E	ST	
Shengeze lake	KZN	8	27 01.37'S; 32 48.82'E	ST	
Kosi Bay river	KZN	6	26 56.90'S; 32 48.96'E	ST	
Nkanini river	KZN	4	26 56.90'S; 32 44.65'E	ST	
Mozambique					
Buzi river		1	19 55.28'S; 32 20.28'E	T	SAIAB67668
Buzi river		1	19 58.39'S; 33 23.34'E	T	SAIAB67656
Changane river		1	22 39.50'S; 33 10 21'E	T	SAIAB75911
***G. tenuiformis*, S. Africa**					
Mzimvubu river	KZN	1	31 34.69'S; 29 29.55'E	ST	
Zinkwazi estuary	KZN	2	29 20.17'S; 31 26.57'E	ST	
Matigulu estuary	KZN	2	29 05.45'S; 31 36.08'E	ST	
Mzimkulu river	KZN	6	28 49.60'S; 31 59.23'E	ST	
***G. giuris*, Swaziland**					
Umbobo		1	26 45.28'S; 31 58.41'E	ST	SAIAB69644
SivungaDam,		1	26 39.41'S; 31 59.03'E	ST	SAIAB70172
Sivunga Dam (East), Usuthu system		1	26 39.40'S; 31 59.10'E	ST	SAIAB70207

Table 2.2. Analysed specimens of other species of *Glossogobius*, provided by other scientist partners from Australia (Art Gallery and Museum of the Northern Territory (NTM), Darwin) and species of outgroup taxa downloaded from Genbank.

Australian	***Glossogobius* species**			
Country	**Locality name**	**Species name**	**Museum numbers/ Genbank number**	**Sample size**
Australia	Daly River, N.T	*Glossogobius* sp.	S.16320-001	1
Australia	Daly River, N.T	*Glossogobius* sp.	S.16484-004	1
Australia	Elizabeth Creek, Daly river	*Glossogobius* sp.	S.16491-009	1
Australia	Sandy Creek, Keep River	*Glossogobius aureus*	S.14778-005	1
Australia	Robinson River	*Glossogobius aureus*	S.16551-001	1
Australia	King River, N.T	*Glossogobius giuris*	S.15865-002	1
Australia	Bradshaw Creek, N.T	*Glossogobius giuris*	S.15866-004	1
Australia	Negiri River, W.A	*Glossogobius giuris*	S.16259-001	3
Australia	Lake Argyle, W.A	*Glossogobius giuris*	S.16408-002	1
Australia	Ord River, W.A	*Glossogobius giuris*	S.16530-001	1
Australia	Ladmilla Creek, N.T	*Psammogobius biocellatus*	S. 16534-001	1
Taiwan		*Glossogobius celebius*	AF265393	2
Belize	Quinaldine	*Bathygobius curacao*	AF391426	1
		Rhinogobius sp.	AY645693.1	1
		Ctenogobius sp.	057622.1	1

Table 2.3. Goby-specific primers used for the amplification of the cytochrome *b* and the nuclear S7 ribosomal protein gene intron 1.

Primer	**Sequence (5' - 3')**	**Source**
Cyt *b*		
GOBYL14673	TAATGGCGTGAAAAACCACCGTTGT	Thacker and Hardman 2005
GOBYH15958	TTTGAGCAG(TAG)AGGGAGATTTTA	Thacker and Hardman 2005
S7 intron1		
S7RPEX1F	TGGCCTCTTCCTTGGCCGTC	Chow and Hazama 1998
S7RPEX2R	GCCTTCAGGTCAGAGTTCAT	Chow and Hazama 1998

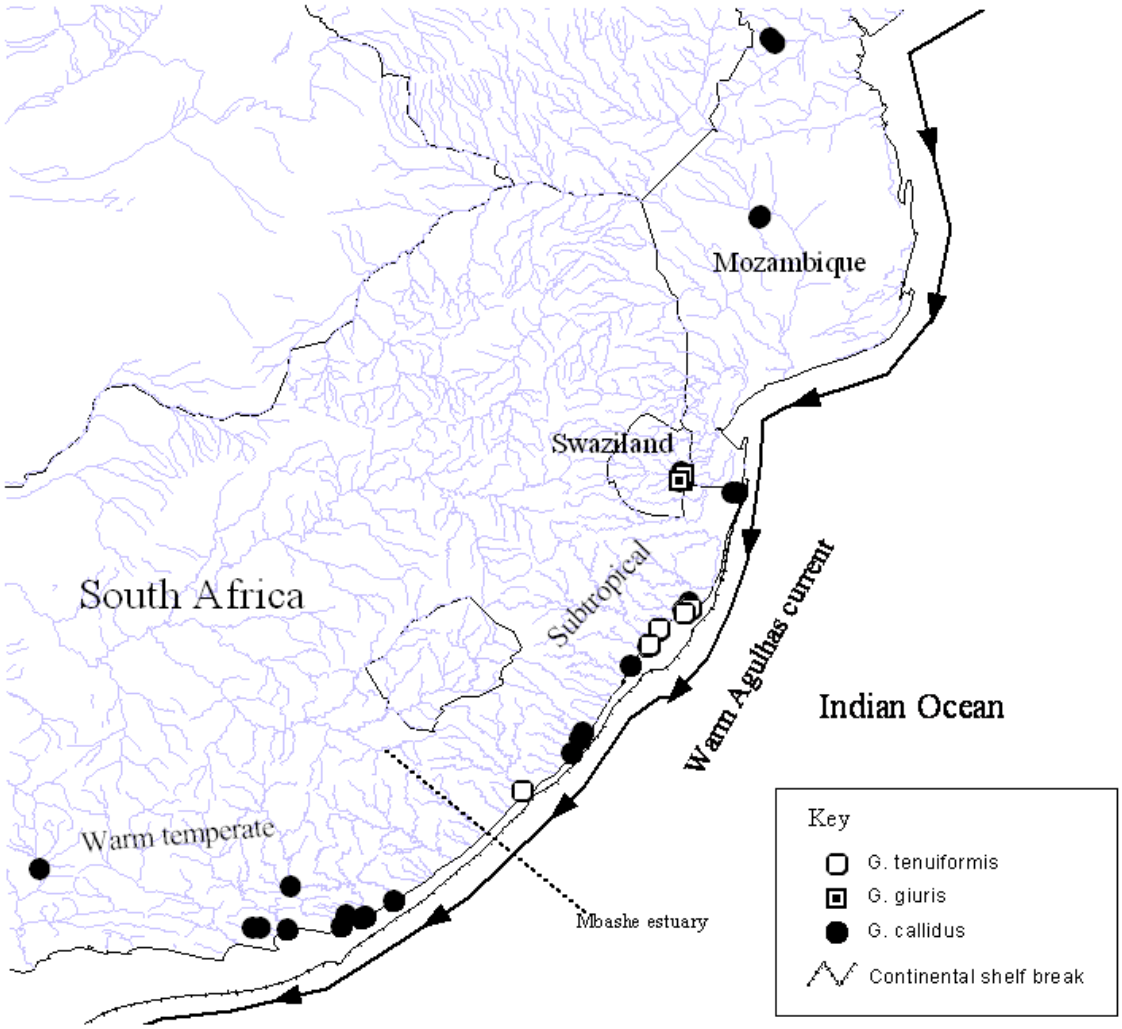


Figure 2.1. Map of southern Africa showing all sampled estuaries and rivers for all three species. The continental shelf break and the influential warm Agulhas Current are also indicated.

CHAPTER 3

POPULATION STRUCTURE AND PHYLOGEOGRAPHY OF *GLOSSOGOBIUS CALLIDUS* SMITH 1937

3. INTRODUCTION

The river goby *Glossogobius callidus* (Figure 1.2), exhibits extensive salinity tolerance as it is found in estuaries as well as freshwater habitats (Whitfield 1998). This species distribution has been recorded to extend from inland Malawi and Mozambique (Tweddle 2007) down to the Swartvlei estuaries of the east coast of the Western Cape Province, South Africa (Figure 1.3 and 2.1) (Whitfield 1998; Skelton 2001). In South Africa, the species also occurs inland to the far northern limits of the Limpopo system in the Limpopo Province (Greenwood 1994; Engelbrecht and Mulder 1998). As previously stated, populations of *G. callidus* are exposed to widely different environments. Therefore selection pressures might favour localised genetic adaptations and, ultimately, genetic divergence (Engelbracht and Mulder 1998). Studies on the biology of this species in both water systems, (Whitfield 1998; Strydom and Neira 2006) have also suggested that adults (and larvae) are one of the dominant groups in terms of biomass.

Although *G. callidus* is common and widely distributed across southern Africa, the phylogeography of this small gobiid species is poorly known. There are probably two reasons that might have caused the lack of attention to genetic structuring of this species are due to the occurrence of the species in both riverine and estuarine systems. Firstly, there is a general hypothesis amongst biologists, that species occurring in estuarine environments have no population structuring (Grant and Bowen 1998). However, several recent studies on fish and other taxa have found that the hypothesis of homogeneous populations among estuarine species is not necessarily true (e.g. Oosthuizen *et al*. 2004; Norton 2006). Secondly, unlike estuarine species, it could be expected that riverine fishes would produce high genetic structuring because barriers to gene flow (such as mountains) may disrupt the link between populations and drainages in terrestrial aquatic systems thereby encouraging genetic diversification to occur over time (Skelton 1980; Stepien and Faber 1998; Swartz *et al*. 2007). Because freshwater fishes are constrained by drainages systems, they have provided insights into the relationships between the contemporary genetic structure and historical changes in the environment (Avise 2000). With the river goby dominantly being found in both freshwater and estuarine systems, studies using molecular markers are of importance in elucidating the

phylogeograpy in these regions to fully understand the patterns and factors that affected their distribution, adaptation and genetic structuring.

To date, allozymes have been the only marker applied to the study of *G. callidus* populations (Engelbrecht and Mulder 1998). Allozymes could potentially provide useful data once valid localised and distant populations have been detected and investigated. Their results tend to show high levels of genetic variation and distinct phylogeographic structure than can be correlated with the allegedly restricted breeding or dispersal capabilities of larvae or adults (Duran *et al*. 2004). Indeed, the results of Engelbrecht and Mulder (1998) have indicated high genetic variation within distant populations of the species. This highlighted the observation of Hoese (1986), based on morphological examination, that the species, *G. callidus*, may represent a complex of cryptic species that may have remained undetected by previous morphological studies. In terms of management and conservation, the above problem may lead to incorrect measures of conservation with detrimental effects on species survival and taxonomy (Moritz 1995). Thus a review of the phylogeography of this species in South Africa will contribute to a better understanding of its cryptic species composition and distribution in our waters, and help delineate areas with management and conservation priorities.

In the present study, a complete mtDNA cytochrome *b* gene and partial nuclear S7 ribosomal protein gene (intron 1) among localised populations of the species were used to: (i) estimate the genetic population structure, (ii) interpret the population genetic differentiation in response to historical drainage systems, and (iii) provide a genetic basis for the conservation of this southern African endemic species, *G. callidus*. The identification of genetically different lineages within the species could also provide a foundation to the taxonomy of the species in response to the observation of cryptic species. The markers used were selected based on their different rates and modes of evolution (see Chapter 1) and because they have, in many studies, shown a clear ability to resolve relationships between and within species, and higher taxa (Akihito *et al*. 2000; Harada *et al*. 2002; Lavoue *et al*. 2003; Coronel *et al*. 2004).

3.1. Materials and methods

All materials and methods are as described in Chapter 2 (Materials and Methods).

3.2. Results

3.2.1 Cytochrome *b* analyses

A preliminary analysis of the nine Eastern Cape (EC) and 12 localities from the KwaZulu-Natal (KZN) region and Mozambique using the cytochrome *b* genetic data revealed that *G. callidus* was composed of two distinct maternal lineages. There was 5% mtDNA sequence divergence between these two lineages compared to the much lower population divergences ranging from as low as 0.10% to 0.87% in the EC lineage and 0.001% to 1.2%, KZN lineage. The data was therefore divided into two groups or datasets which were recognised as the EC and KZN lineages for all subsequent intraspecific analyses (Table 3.1). The widespread EC lineage (estuarine/freshwater lineage) consisted of all of the sampled localities from the EC as well as seven sampled localities from KZN resulting in a total of 131 individuals for analysis. The second lineage, KZN lineage (a predominantly freshwater lineage), comprised of five localities from KZN region, and a few individuals collected in Keisekamma estuary in the Eastern Cape region. The KZN lineage was therefore smaller and restricted in comparison, consisting of only seven localities and 39 individuals for analyses (Table 3.2).

3.2.1.1 Genetic diversity and haplotype relationships

The EC lineage sequences consisted of 1064 nucleotide base pairs of which 72 were polymorphic/ variable while 992 were constant or invariable sites. Of the polymorphic sites 47 were parsimony informative while 25 were singletons and thus not useful in the analyses. A total of 56 unique haplotypes (Table 3.3) were estimated in DnaSP with a very high haplotype diversity (h = 0.92) but low nucleotide diversity (π = 0.004) per site. The observed haplotype diversity was at its lowest in the West Kleinemonde (h = 0.22) and high at Mpenjati and Mzimayi localities (h = 1.00). Nucleotide diversity in this lineage was low in the Great Fish and West Kleinemonde (π = 0.00) and moderate in Mzingazi lake (π = 0.04) (Table 3.2). The average number of pairwise differences observed was relatively low at 4.28.

The most widely distributed or commonly shared haplotypes in the EC lineage were haplotype H7 and haplotype H1 (Table 3.3 and Figure 3.1a). Haplotype 7 was represented by 31 individuals (24% of analysed individuals) from seven warm-temperate localities (SN, GB, KA, WK, EK, KK and KO) and four subtropical localities (MM, MT, MLZ, and ZO) (Tables 3.1 and 3.3 Figure 3.1). H1 though common was more restricted and only distributed among the localities of the Eastern Cape (i.e. GB,

KK, KA, and GF). Of the 17 haplotypes that were shared among individuals, only nine were distributed in more than one locality (H1 and H7 mentioned above plus H4, H6, H10, H11, H14, H29 and H34). There were also a number of haplotypes (39) that were rare and not shared among localities. The network of haplotypes revealed four geographically distinct groups within the EC lineage, namely: From warm-temperate, BU-SN (Bushman's + Sundays), grey haplotypes dominated group, GF-KK-E/WK-GB (Great Fish, Keisekamma, East/West Kleinemonde and Gonubie), blue haplotypes dominated group, KA-KO (Kariega + Kowie), brown haplotypes dominated group, and subtropical, NK-KS (Nkanini and Kosi Bay), green haplotypes dominated group (Figure 3.1a). There were only two missing variables (mV) on the network of haplotypes which are indicative of representative population coverage in the study. These missing variables represent the missing haplotypes that may have not been sampled during the course of the study.

Genetic structuring was more evident among haplotypes from the localities sampled in the warm-temperate Eastern Cape, while those sampled in the subtropical KwaZulu-Natal, with the exception of Nkanini and Kosi Bay, were randomly scattered around common haplotype H7, forming a star-like phylogeny (Figure 3.1a). It has been observed that surviving ancestral haplotypes generally serve as internal nodes that have differentiated into the descendant haplotypes. These internal nodes generally have the greatest degree of connectivity, are widely distributed and hence have the highest probability of being the oldest in the network (Templeton *et al*. 1995). Therefore, the haplotypes H7 have the highest probability of being ancestral, or older, for the EC lineage dataset of *G. callidus,* followed by haplotype H1.

Analysis of genetic diversity in the KZN lineage consisted of 39 sequences which had 48 variable sites and 1016 invariable sites in total. Only 39 characters were parsimony informative while nine were parsimony uninformative or singletons. Sequence divergence between localities within this lineage was low in comparison to the EC lineage with the exception of rare Mozambique haplotypes. A total of 18 haplotypes (Table 3.3 and Figure 3.1b) were estimated from the 39 analysed individuals (h = 0.91). Similar to the EC lineage, nucleotide diversity was also low (π = 0.005), with observed pairwise differences of 5.33. With the exception of the Mozambique population's low nucleotide diversity, Shengeze Lake was the lowest with π = 0.0004 and h = 0.43 diversities (Table 3.2). As per Table 3.3 and Figure 3.1b, the most common haplotype, with a probability of being an ancestral haplotype, in this lineage was H11 which was represented by nine individuals (29%). Only seven haplotypes were shared among individuals with only four (H5, H6, H11 and H13) found in more than one locality. All other haplotypes were rare and only represented by one individual. The

network of haplotypes (Figure 3.1b) within this lineage did not reveal high geographic structuring as observed in the EC lineage with only haplotypes collected from Mdloti River (H1 – H4) and Shengeze Lake (H7 and H8) being geographically structured.

3.2.1.2 Population demography analyses

The mismatch distribution of the EC lineages was unimodal suggesting a model of a sudden population expansion (Figure 3.2a). The KZN lineage was however bimodal indicating two populations with high genetic differentiation (Figure 3.2b). The calculated raggedness index (*r*) (*r* = 0.006) for the EC lineage was relatively low, while that of the KZN lineage was high at 0.033. A lower value of *r* is indicative of a "smooth" mismatch distribution which in turn could suggest population expansion in a species and the higher value indicates population fragmentation. The pairwise sequence analysis for mismatch distribution (pairwise mean = 4.28) of the EC lineage indicated that some sequence comparisons were very high and suggested a small peak for those individuals, possibly reflecting the geographic structuring observed in other analyses (Figure 3.2a). In the KZN lineage the pairwise differences were relatively high (5.33) and in agreement with the distinct double peaks that were observed (Figure 3.2b), indicating possible an expanding South African population and the low representing the Mozambique population. These findings of a sudden population expansion for the EC lineages are strongly supported by the Fu's *Fs* and the Tajima's *D* tests of neutrality. The observed high and significant negative F_S value (-33.89 and -4.20 for EC and KZN respectively) and the negative Tajima's *D* statistic (-2.15 and -1.93, EC and KZN respectively) in both lineages indicate strong support for the population expansion model.

3.2.1.3 Within-lineage population structure, AMOVA

Within both *G. callidus* lineages, the estimated components of genetic structure and gene flow among the localities that were inferred from the dispersal parameters in DnaSP were very low (average F_{ST} = 0.27, Nm (number of migrants) = 0.65 in EC lineage; average F_{ST} = 0.83, Nm = 0.05 in the KZN lineage). Although F_{ST} in the EC lineage seems very low, the network of haplotypes revealed isolation of geographic lineages, while the Nm values also suggest low migration and thus isolation of localised populations. Further analyses of genetic structuring and subdivision of the pairwise Φ_{ST} values (Table 3.4) among the localities using the Tamura-Nei model of substitution with the gamma correction (0.19) found in ModelTest were also high.

In the EC lineage, most pairwise Φ_{ST} locality comparisons were not statistically significant. However, significant pairwise Φ_{ST} tests were observed when all localities were compared with the

Nkanini (NK), Kosi Bay (KS) or Mzimayi (MM) localities from the KZN as well as the Great Fish (GF) from the EC region (Table 3.4). It should however be noted that the Great Fish estuary was probably unique in that all ten individuals analysed from this river shared the same haplotype in the sample (h = 0.0). Because of the isolation of this population, the single haplotype may be an indication of a young population as a result of recent invasion. Pairwise Φ_{ST} comparisons including the (KK) and Bushman's (BU) estuaries were also high but not significant. This may be due to the absence of the widespread ancestral haplotype among the individuals of these localities. An analysis of molecular variance estimated indicated that structuring within this EC lineage could be expected as the pairwise genetic variation between all localities is estimated at Φ_{ST} = 0.48. AMOVA further indicated that 51.58% variation was found within localities, while a significant 38.66% of the variation (Φ_{CT} = 0.39) was observed among the three defined groups (Table 3.5). Results of AMOVA in which localities were divided into three groups (1) Nkanini + Kosi Bay, (2) Mzimayi locality alone, (3) Bushman's estuary (type locality and the rest of the localities) indicated that very little of the variation (9.76%) could be attributed to variation among localities within groups (Table 3.5). The results indicate that the significant genetic variation in this group or 'species' was explained at the level of within localities and among the defined geographic groupings.

In the KZN lineage, pairwise variation (Φ_{ST}) between localities was in general higher and more significant than in the EC lineage (Table 3.6). The significant comparisons were observed when individuals from Mozambique (MQ), Shengeze Lake (SH) and Mdloti (MD) were compared with other locations. This indicates some genetic subdivision and variation between all localities and these three (MQ, SH, and MD) localities. An AMOVA indicated that higher percentage of variation (73.7%) in this lineage was observed among the three maximized groups (Φ_{CT}= 0.73) (i.e. (1) Mozambique; Mdloti and Shengeze Lake, (2) Mzingazi, Mhlatuzi, Mzumvubu and (3) Keisekamma) (Table 3.7)). Results further indicated that very little percentage of the variation (9.37%) could be attributed to variation among localities within each groups (Φ_{SC}= 0.35). The results indicate that the significant genetic variation in this lineage (Φ_{ST}= 0.83) was relatively well explained both at the level of among groups and within localities.

3.2.1.4 Phylogenetic tree reconstruction

To calculate the support for the phylogeographic structuring observed in the network of haplotypes and AMOVA, a phylogenetic tree was reconstructed from the unique haplotypes by using maximum likelihood (ML) only. ModelTest (Posada and Crandall 1998) was used to select the best model for the data for the maximum likelihood analyses. The base frequencies of the sequences were estimated

as 0.24% (A), 0.36% (C), 0.15% (G) and 0.25% (T), while the shape parameter of the gamma distribution was 0.19. The best model of evolution for the data was based on the Akaike Information Criterion (AIC) and estimated as the TrN+G model. The parameters obtained from this analysis and from the model were used for the construction of the ML tree.

The estimated maximum likelihood tree (Figure 3.3) (–lnL 3776.07) supported the same maternal lineages (EC and KZN lineages (100% bootstrap), as discussed above) and geographic structuring as observed in both (a) and (b) in Figure 3.1. Even though the two lineages are genetically distinct, they could not be differentiated by morphological assessment using both light and scanning electron microscopy (SEM) (see Figure 3.6). The two *G. callidus* individuals from inland freshwater localities of Mozambique were part of the KZN lineage but as a unique monophyletic clade with a genetic divergence of 1.2%, suggesting an additionally more divergent lineage of the species. Even though some groups (i.e. A to C, F) had low bootstrap support (<60%) the tree supported the same phylogeographic structuring depicted by the network of haplotypes and AMOVA. For the EC lineage groups A (Great Fish, Kleinemond, Gonubie and Keisekamma), B (Bushmans and Sundays), C (Kariega and Kowie) and clade D (Kosi Bay and Nkanini) (98% bootstrap) could be distinguished. (Figure 3.3) In the KZN lineage, only few localities were phylogeographically structured, namely the only the two haplotypes collected from Shengeze Lake (group F) and the divergent Mozambique population (clade E, 100% bootstrap) (Figure 3.3). The monophyly of the Mdloti haplotypes (H1 to H4), (Figure 3.1b) was not supported by this phylogenetic analysis.

3.2.2 Nuclear S7 ribosomal protein gene intron 1

3.2.2.1 Genetic diversity and haplotype relationships

The sequence divergence was estimated as 1.5% between the two lineages (EC and KZN lineages as previously estimated by the cyt*b* data) for the S7 data set. Consequently, the data were then analysed as a single population. Due to difficulties experienced in the amplification and sequencing of the S7 intron 1 gene, the sample size was smaller and reduced to a total of 102 sequences (857 base pairs long) from 18 localities (Table 3.8). These sequences had 41 polymorphic/ variable sites and 816 invariable sites. Of the polymorphic sites, 28 were parsimony informative while 13 sites were singletons and thus not useful in the analysis. The highest DNA divergence was observed between Mhlatuzi (UM) and Keisekamma (KK) (0.31%) while the lowest was observed between the Kowie (KO) and Sundays River (SN) (0.01%). A total of 49 unique haplotypes (Table 3.9 and Figure 3.4) were estimated with a very high haplotype diversity (h = 0.95) but low nucleotide diversity (π =

0.009) per site. The observed haplotype and nucleotide diversity were the lowest in Mzimvubu (h = 0.00, π = 0.00) due to the low sample size (Table 3.8). Other localities with low nucleotide diversity were the Great Fish (GF), Mdloti (MD), Mpenjati (MP) and Kowie (KO) systems (Table 3.8). The average number of pairwise differences observed in the sample was relatively high at 7.89.

As indicated in Table 3.9, the most commonly shared haplotype(s) in this data set are haplotype H1, H2, H21 and H20. Haplotype 1 was the most common with 17% of the analysed individuals representing five warm-temperate localities (KE, KA, GF, KK, and KO) (Table 3.8, 3.9 and Figure 3.4). Haplotype 21 was restricted and only shared among subtropical localities (MLZ, MP and MT). Of all the haplotypes, 13 haplotypes were shared among individuals with ten of these being distributed in more than one locality (H1, H2, H5, H15, H17-21, and H24). The minimum spanning network of S7 haplotypes also revealed two diverged clades/lineages (Figure 3.4, A and B), with a maximum of nine mutational differences between them. These lineages consisted of the same localities and restriction as was discovered in section 3.2.1 in the preliminary analysis of the cyt *b*. The most restricted lineage was that recognised as the KZN lineage (Figure 3.4 A, more subtropical lineage) in section 3.2.1 (cytochrome *b*), whilst the most widely distributed lineage, overlapping with the KZN lineage, was the EC lineage (Figure 3.4 B, more warm-temperate lineage). In the latter lineage, haplotypes collected in the subtropical localities formed monophyletic groupings, i.e. haplotypes collected from the far north of KZN (Kosi Bay+Nkanini) were geographically structured (H17, H18 and H46), while all other localities south of Kosi Bay (MLZ, MP, MT and MM) were also monophyletic (H21, H24 – 32), (Figure 3.4). Among the six localities that formed the KZN lineage (Figure 3.4 A), most of the haplotypes from Mhlatuzi (H20, and H41 – 45) were restricted to the same part of the network. All the haplotypes collected from Keisekamma (H6 – H11 and H 15) belonging to this lineage were also geographically monophyletic. The two haplotypes collected from Mozambique (H20 and H49) were also found within the KZN lineage, with one (H20) common to Mhlatuzi and Mdloti localities.

3.2.2.2 Population demography analyses

The mismatch distribution analysis revealed a bimodal distribution (Figure 3.5), with each distribution representing possibly each lineage, and suggesting a model of a sudden population expansion for each distribution. The calculated raggedness index (r) was low (r = 0.019), a value indicating a “smooth” mismatch distribution which in turn could also suggest population expansion in a species. The pairwise sequence analysis for mismatch distribution (pairwise mean = 7.89) was very high suggesting high divergences between the two observed peaks or lineages. These findings

of a sudden population expansion were also supported by the Fu's F_S and the Tajima's *D* tests of neutrality. The observed significantly negative F_S value (-25.11) and the negative Tajima's *D* statistic (-0.21) strongly indicate a population expansion model. This was further supported by the observed high haplotype and low nucleotide diversities.

3.2.2.3 Within-lineage population structure, AMOVA

In a population with high F_{ST} indices, the numbers of migrants per generation tend to be relatively low. Indeed a high $F_{ST} = 0.69$ was observed, with a low number of migrants (Nm = 0.11) exchanged between the two lineages, and within lineages. These results together with the network of haplotypes (Figure 3.4) revealed separation and structuring within this species.

Further analyses of genetic structuring and subdivision were carried out calculating pairwise Φ_{ST} values among all localities using the Kimura 2P model of substitution as the closest model to the one selected by AIC in Modeltest (Posada and Crandal 1998). Pairwise Φ_{ST} comparisons were statistically significant between localities found in different lineages (Table 3.10). Low pairwise Φ_{ST} values were observed when localities within one lineage were compared. Some localities exhibited a negative Φ_{ST} value between their pairwise comparisons which statistically implies that genetic comparisons between those localities were too low to be detected. Pairwise Φ_{ST} comparisons including Mzumvubu, Mhlatuzi, Mdloti, Shengeze and Mozambique (Subtropical localities) were the highest compared to the localities in the warm-temperate EC region.

An analysis of molecular variance (AMOVA) estimated that a significant percentage of variation (65.6%) was observed between two groups ($\Phi_{CT} = 0.66$), namely EC lineage and KZN lineage (see Table 3.1), while 25.22% of the variation ($\Phi_{SC} = 0.27$) was observed within the groups (Table 3.11). Results also revealed high Φst (0.74) between the two groups, while indicating that a small part of the variation (9.15%) could be attributed to variation among localities within each group. Furthermore, the significant genetic variation in this species was explained at the level of among groups and within localities as observed with the cyt *b* dataset.

3.3. Discussion and conclusion

Investigation of large-scale phylogeographic patterns of the southern Africa endemic *Glossogobius callidus* revealed that the species is divided into two highly divergent lineages in both used genetic markers (EC and KZN lineages) that are associated with the coastal biogeographic provinces (Figure 2.1). The separation of the species into two lineages corresponds to results by Engelbrecht and Mulder (1998) (in the allozyme study) suggesting that the species is genetically sub-divided. As can be observed in Table 3.1., most analysed localities within the EC lineage are estuarine and permanently open systems; while those in the KZN lineage are all freshwater riverine systems that could have been isolated for a long time. Although both genes differentiated this species (*G. callidus*) into these two clades, specimens could not be separated by the SEM morphological assessment of papillae patterns as only one pattern of cheek papillae similar to the ones observed by Helen Larson and associates (*pers. comm.*) was observed (Figure 3.6). This could be due to the fact that diagnostic morphological characters for *Glossogobius* species are known to be very limited and can be expected to be few between closely related or recently diverged lineages or species (Akihito and Meguro 1975; Hoese and Winterbottom 1979; Penzo *et al*. 1998; Pusey *et al*. 2004). Although morphologically unconfirmed at this point, the results of this study strongly suggested that *G. callidus* consists cryptic species (according to phylogenetic species concept) in southern Africa that are morphologically very similar, but genetically distinct.

Cytochrome *b* data indicated that the effects of river system history and or biogeography structure on phylogeographic parameters may differ considerably within each of these lineages. In the EC lineage, adjacent localities found only in the warm-temperate region were highly structured phylogeographically, providing evidence of regional population fragmentation and isolation, while all localities found in the subtropical region, with the exception of Kosi Bay and Nkanini, were genetically homogeneous with high levels of gene flow between them. The observed population structuring or geographic lineages within the EC lineage is apparently associated with riverine localities (see Table 3.1), while estuarine localities are not geographically structured. This is relevant since the system types of these localities (especially in the subtropical region) are classified as temporary open systems. It can thus be hypothesised that when the estuarine system floods, the individuals of the species are flushed to the sea such that the Agulhas Current may influence the movement and recruitment of the individuals into new localities. Although marine dispersal of the studied species has not been recorded, the possibility of sea dispersal cannot be ruled out as it has been reported for other species of the genus such as *Glossogobius giuris* (see Islam 2004). This author reported that even though the species is found mainly in freshwater and estuaries, the species

can also enter the sea. Although the results for the KZN lineage are inconclusive as only a few individuals were analysed from this group, these results did indicate a possible species homogeneity or gene flow between its localities and between the two coastal biogeographic regions for this population. The low levels of phylogeographic structuring observed in the subtropical region of the EC lineage (Figures 3.1 and 3.3) might not be due to low overall genetic diversity, but rather to limited time for differentiation of nucleotides in the cyt *b* gene as estimated in the high haplotype diversities but low nucleotide diversities (see Table 3.2) but could rather be caused by the small number of mutations observed in the cyt *b* gene data that could not be resolved at the intraspecific level.

The estimated phylogenetic tree in the present study further indicated that although individuals from different localities do not form reciprocal monophyly, localities from adjacent systems in the warm-temperate region (EC lineage) cluster geographically and had significant Φ_{ST} pairwise comparisons. For example, all the riverine localities south of Great Fish River (BU-SN) were within group B, while all localities north of this river (KK, KD, GB and GF) clustered together within group A. The majority of individuals from the adjacent Kariega and Kowie localities (group C) also formed a separate monophyletic group (51% bootstrap support) that was differentiated from localities west of the Great Fish estuary. Only a few haplotypes from these localities, Kariega and Kowie, were found in moderate numbers in both groups A and B (Figure 3.3). Of the three groups identified above, the most EC region restricted group was group B (BU, SN) as it was composed of all the Sunday's and 60% of Bushman's river haplotypes. Skelton (1980b) reported that large drainages of the Eastern Cape, such as the Great Fish and Bushman's, would have expanded naturally during sea level changes in the past, coming in contact with adjacent systems, making it possible for fish to disperse among them. In addition, Whitfield (1998) reported that *G. callidus* in the Eastern Cape breeds in the upper regions of the river-estuary where individuals are at low risk of being flushed to the sea. The majority of the localities in the EC classified as permanently open systems may reduce the risk of individuals being flushed to the sea since the water flow to the marine environment is stable and always at a moderate flow. Both the phenomenon of sea level changes and the river capture activities and the above mentioned factors could have however contributed to the structuring of adjacent rivers and the low variation observed within them today as well as the pattern of isolation observed. The fact that the entire observed EC geographic groups shares a common possible ancestral haplotype (H7), supported a historical connection between the analysed specimens and adjacent rivers in this region. It is important to note that same pattern of structuring was also observed with the nuclear marker (S7 intron 1) in the EC and KZN biogeographic regions. Quattro *et al*. (2002) suggested that

a greater depth in Φ_{ST} in AMOVA analyses indicates a phylogenetic component with regard to the separation of populations or species. The results in this study therefore reflect a pattern consistent with this hypothesis, and reflect a phylogenetic distinctiveness and structuring of few localities.

With the exception of the two 5% diverged *G. callidus* lineages (EC and KZN), the other most diverged geographic lineage (with a 1.2% divergence, cyt *b*, 100% bootstrap support) was the inland freshwater Mozambique lineage, clade E (Figure 3.3). This clade was part of the KZN lineage and could represent a distinct Evolutionarily Significant Unit (ESU) or Management Unit (MU). This inland freshwater clade has been isolated from the main species range and may be reproductively isolated as well. Consequently, it is also of importance for understanding the taxonomy of the *G. callidus* group. However, more sampling is required in this area to support this conclusion as only a few specimens were available for analyses. The results also indicated another diverged freshwater lineage of the species in the EC lineage. The Nkanini and Kosi Bay lineage was about 0.87% divergent (cyt *b*) from the other localities within the EC lineage. This together with the somewhat high divergence suggests that the Nkanini and Kosi Bay may have been formed due to founder effects or after successful migration into new areas as a result of past river system connections with the widespread *G. callidus* form. These localities could not be linked to the KZN lineage as was expected, as they belonged to the widespread EC lineage. The Nkanini and Kosi Bay areas are geographically isolated from the main part of the species range and could be in the process of becoming reproductively isolated, thus the high genetic divergence. However, isolation may have occurred only too recently for their populations to have attained complete isolation from the rest of the species range. Alternatively, although these rivers may have been geographically isolated in the past and reached reciprocal isolation, they could currently be receiving immigrants from the other parts of the wider species' distribution range in order to maintain gene flow while avoiding speciation within the lineage. The combination of all results from both analysed markers (cyt *b* and S7 intron 1) support the first scenario based on the observation that all localities (with the exception of the close locality Lake Mzingazi) do not share haplotypes with these two systems (Nkanini and Kosi Bay). There is also no record of previous connections between the two and the other analysed localities. The observed pairwise sequence comparisons of these two localities were also significantly high in both genes (Tables 3.4 and 3.10), thus further suggesting isolation and no migration of individuals between the two rivers (Nkanini and Kosi Bay) and the rest of the species' populations.

On the other hand, the results also suggested a genetic link between the freshwater South African *G. callidus* specimens from Shengeze Lake (clade F, Figure 3.3, KZN lineage) and the individuals from the Buzi and Changane rivers in Mozambique (group E, Figure 3.3 and 3.4). Although Shengeze Lake and the Kosi Bay system are geographically closer (less than 10 km apart), they were found to belong to the identified two distinct lineages of *G. callidus*. While the specimens from the Kosi Bay system were part of the widespread EC lineage, the Shengeze lake haplotypes were well supported as a monophyletic clade and represented only within the KZN lineage. These genetic analyses therefore indicate that past historical processes may be very important in the observed geographical structuring of *G. callidus.* Shengeze Lake is a small freshwater system which is geographically isolated from other systems in the region. This isolation might have been caused during sea level changes when river capture between drainages were disrupted resulting in the isolation of Shengeze Lake. Founder effects or possible contiguous range expansion followed by extended periods of isolation (fragmentation) and restricted gene flow appear to be the main mechanisms for the increased variation and structuring within the species (see Figure 3.1 and 3.4) between these localities. The genetic separation between both Shengeze Lake and Kosi Bay localities can also be linked to the structure and formation of the Kosi Bay system within southern Maputaland, which is estimated to have formed during the middle to late Pleistocene (Momade and Achimo 2003; Porat and Botha 2008). Pre-adaptation and continued expansion into suitable habitats could have allowed the species to successfully colonise new areas (such as Mozambique or vice versa), regardless of population size or genetic diversity of the founder populations. Indeed the estimated divergence rates as estimated for the cytochrome *b* gene (5%) support a Pleistocene expansion using the conserved cyt *b* molecular clock of 2% variation per million years (MYA) (Kocher and Stepien 1997).

Furthermore, southern Maputaland stretches from the border of Mozambique down to Sodwana Bay, South Africa (Momade and Achimo 2003) and is characterised by gradients in coastal environmental conditions and changes during the Pleistocene that in combination may represent a past and present formidable barrier to the dispersal of coastal organisms. Southern Maputaland has been identified as a transition zone between tropical Indo-Pacific and subtropical southeast African biota (Stuckenberg 1969), extending down to and ending at Cape Vidal (Sink *et al.* 2005). In this area, the continental shelf widens, the influence of the warm southward-flowing Agulhas Current is reduced, and the coastal environment is dominated by a persistent upwelling cell (Schumann 1988). The scarcity of suitable habitats between Kosi Bay and Richards Bay, caused by possible barriers such as dunes, may limit the amount of gene flow between populations of some coastal and inland riverine species (Flemming and Hay 1988) e.g. Nkanini, Shengeze and Kosi Bay populations. The same pattern of

separation, for example, was also observed in the cumacean *Iphinoe truncata* Hale, 1953 in which samples from the Kosi Bay system were differentiated from other populations from the east and southeast coast of South Africa (Teske *et al*. 2007b). These patterns of phylogeography have also been observed in other taxa where populations from this region formed monophyletic groupings with haplotypes north of the region in Mozambique (Sink *et al.* 2005; Teske *et al*. 2009). However, in this present study, the Kosi Bay and Nkanini haplotypes were isolated from the lineage of the freshwater Mozambique individuals, i.e. within the KZN lineage. The factors shaping the above discussed structure might have formed a possible phylogeographic break separating the Kosi Bay and Nkanini haplotypes, which are observed to be significantly divergent from the rest of the localities in the EC lineage.

Estimation of population expansions and demography requires several assumptions; hence, we evaluated several methods, including Tajima's *D*, Fu's F_S, and mismatch distributions. Overall, in contrast to the high haplotype diversity and lower nucleotide diversity, the population demography (Figure 3.3 and 3.5) suggests a bottleneck followed by population expansion (Grant and Bowen 1998), which is generally associated with little or no population structure (von der Heyden *et al*. 2007). The neutrality test of mtDNA cyt *b* and nuclear S7 intron 1 was rejected based on Tajima's *D* and Fu's tests. Both tests are sensitive to factors such as bottlenecks and population expansion, which would tend to drive the values towards being more negative (Tajima 1989). Indeed, significant negative values of the F_S (-33.887 and -10.285, for EC and KZN respectively), as well as *D* (-2.15 and -1.69, for EC and KZN respectively) indicated that *G. callidus* in South Africa had experienced historically rapid population expansion. The very same patterns of significant Fu F_S (-25.11) and Tajima's D (-0.211) values were also revealed by the S7 intron 1 gene. In addition, the mismatch distribution (Figure 3.2 and 3.5) supported the model of expansion and sequential colonisation of adjacent river systems within this species. In the EC lineage, however, various peaks were observed (Figure 3.2a), which might suggest that fragmentation, had played a role in shaping the genetic variation and structuring found in this lineage. The bimodal mismatch distribution observed for the S7 intron 1 (Figure 3.5) also further provided strong evidence for the presence of two genetically distinct lineages of the species. These are reproductively and genetically isolated (5% divergent) and may thus represent unique species.

The assumption made by the mismatch analyses, as well as the low raggedness statistics (r = 0.006, cyt *b*, and r = 0.019, S7 intron1) was that the observed genetic structure is due to a single major ancient expansion. However, it seems improbable that the complex genetic structuring found within

the EC lineage could be explained by a single expansion event. As observed in the network of haplotypes (Figure 3.1a and 3.4) and confirmed by the phylogenetic tree (Figure 3.3), it is more likely the result of small-scale expansions into suitable habitats and adjacent river systems over time. Sea level changes, regressions and transgressions, during glacial cycles would have resulted in the connections and separations of river systems, especially within the south coastal localities (Skelton 1980a, b; Swartz *et al*. 2007). These events would have made fish movements and hence gene flow possible between the different systems at different times. These changes could also have resulted in the processes of isolation or fragmentation, during periods when rivers and estuaries sharing confluences were separated or had become independent from each other. The signatures of such processes are evident in the high haplotype and nucleotide diversity values, the ragged mismatch distribution, and the levels of structuring, isolation of some localities and high pairwise differences. Additionally populations of several other gobiid fishes in South African waters (Neethling *et al*. 2008) were also characterised by unimodal distributions with low pairwise differences, indicating newly expanded populations. For the river goby populations, however, the observed pairwise differences were large (mean of 7.86 for S7 intron 1, and 4.28 in EC lineage), corresponding well with an ancient/old population expansion, as also reported and observed among other gobies (*Clevelandias ios* Jordan and Gilbert, 1882) (Dawson *et al*. 2002).

In conclusion, *G. callidus* is a species with strong phylogeographic structure compared to other studied goby species in South Africa (such as Neethling *et al*. 2008, Mavimbela, MSc thesis in progress). Factors affecting the species genetic structuring within each of the biogeographic regions are different as the processes of isolation and connectivity were different. Within the EC lineage, the subtropical province populations have a high level of gene flow and thus low genetic structuring, whereas in the warm-temperate province there is high genetic structure due to low levels of gene flow between localised populations. This structuring is associated with the formation of past river capture that connected localities in the nearest proximity. The pattern of structuring observed in this biogeographic region is also either as a result of a vicariant event, such as the sea level changes, that caused the colonisation and the separation of the rivers and estuaries via the sea, or a combination of the two phenomena. The results clearly indicated that the species is genetically divergent with a restricted freshwater group (KZN lineage) and a more widespread estuarine group (EC lineage). In connection to the hypothesis of a structured freshwater species, *Glossogobius callidus* individuals collected from riverine water type within the EC lineage are geographically well structured, while the estuarine individuals are less structured, as expected.

Table 3.1. The total number of sampled localities, and the groupings of the localities into different provinces. The type of water habitat (riverine or estuarine), province (EC Eastern Cape and KZN KwaZulu-Natal), and mouth type (PO - permanently open and TO - temporary open) are also indicated. Localities are assigned abbreviation codes that are used in the text.

Locality name	Code in text	SA biogeographical region	Water type	Mouth type	Province
Sundays	SN	Warm temperate	Riverine	Estuarine PO	EC
Bushman's	BU	Warm temperate	Estuarine	Estuarine PO	EC
Kariega	KA	Warm temperate	Riverine	Estuarine PO	EC
Kowie	KO	Warm temperate	Estuarine	Estuarine PO	EC
Great Fish	GF	Warm temperate	Riverine	Estuarine PO	EC
West Kleinemonde	WK	Warm temperate	Estuarine	Estuarine TO	EC
East Kleinemonde	EK	Warm temperate	Estuarine	Estuarine TO	EC
Gonubie	GB	Warm temperate	Estuarine	Estuarine PO	EC
Keisekamma	KK	Warm temperate	Riverine	Estuarine PO	EC
Mzumvubu	UZ	Subtropical	Riverine	Estuarine TO	KZN
Mpenjati	MP	Subtropical	Estuarine	Estuarine TO	KZN
Zotcha	ZO	Subtropical	Estuarine	Estuarine TO	KZN
Mtentweni	MT	Subtropical	Estuarine	Estuarine TO	KZN
Mzimayi	MM	Subtropical	Estuarine	Estuarine TO	KZN
Mdloti	MD	Subtropical	Riverine	Estuarine TO	KZN
Mhlatuzi	UM	Subtropical	Estuarine	Estuarine bay	KZN
Mzingazi Lake	MLZ	Subtropical	Riverine	Estuarine TO	KZN
Mzingazi canal	MZ	Subtropical	Riverine	Estuarine TO	KZN
Kosi Bay	KS	Subtropical	Riverine	Estuarine lake	KZN
Nkanini	NK	Subtropical	Riverine	Riverine lake	KZN
Shengeze lake	SH	Subtropical	Riverine	Riverine lake	KZN
Mozambique	MQ	Tropical	Riverine	-	MQ

Table 3.2. List of localities and their respective lineages and biogeographic provinces, with the number of sequences, haplotypes per locality as well as the haplotype and nucleotide diversity estimates from the cytochrome *b* data. Locality abbreviations are as indicated in Table 3.1.

Locality	Number of sequences	Number of haplotypes	Haplotype diversity (*h*)	Nucleotide diversity (π)
EC lineage				
		Warm-Temperate		
SN	9	5	0.86	0.0034
BU	10	6	0.89	0.0040
KA	16	10	0.94	0.0040
KO	12	8	0.89	0.0036
GF	10	1	0.00	0.0000
WK	9	2	0.22	0.0006
EK	8	5	0.79	0.0027
GB	6	3	0.73	0.0018
KK	7	6	0.95	0.0028
		Subtopical localities		
MP	7	7	1.00	0.0041
ZO	5	4	0.90	0.0017
MT	11	9	0.96	0.0022
MM	5	5	1.00	0.0083
MLZ	6	4	0.86	0.0405
KS	6	4	0.80	0.0053
NK	4	2	0.50	0.0033
KZN lineage				
		Subtopical localities		
KK	8	5	0.85	0.0019
UZ	4	4	1.00	0.0028
MD	5	4	0.90	0.0015
UM	10	6	0.77	0.0012
MZ	2	2	1.00	0.0019
		Tropical		
SH	8	2	0.42	0.0004
MQ	2	1	0.00	0.0000

Table 3.3. Distribution of haplotype diversity from mitochondrial cyt *b* sequences of *Glossogobius callidus* specimens among localities. Locality abbreviations are according to Table 3.1. Haplotypes with a high number of representatives and possibility of being ancestral are marked with a *.

A. Haplotypes from EC lineage

	SN	BU	KA	KO	GF	WK	EK	GB	KK	MP	ZO	MT	MM	MLZ	KS	NK	Total
H1*	2	2	2	_	10	_	_	2	2	_	_	_	_	_	_	_	20
H2	_	3	_	_	_	_	_	_	_	_	_	_	_	_	_	_	3
H3	_	2	_	_	_	_	_	_	_	_	_	_	_	_	_	_	2
H3	_	1	_	_	_	_	_	_	_	_	_	_	_	_	_	_	1
H4	_	1	3	_	_	_	_	_	_	_	_	_	_	_	_	_	4
H5	_	1	_	_	_	_	_	_	_	_	_	_	_	_	_	_	1
H6	_	_	1	_	_	_	_	_	_	_	_	_	_	_	_	_	1
H7*	2	_	2	4	_	8	4	3	1	_	2	2	1	2	_	_	31
H8	_	_	_	_	_	_	1	_	_	_	_	_	_	_	_	_	1
H9	_	_	_	_	_	_	1	_	_	_	_	_	_	_	_	_	1
H10	_	_	2	1			1	_	_	_	_	_	_	_	_	_	4
H11	_	_	_	2	_	_	1	_	_	_	_	_	_	_	_	_	3
H12	_	_	_	_	_	_	_	1	_	_	_	_	_	_	_	_	1
H13	_	_	2	_	_		_	_	_	_	_	_	_	_	_	_	2
H14	_	_	1	_	_	1	_	_	_	_	_	_	_	_	_	_	2
H15	_	_	1	_	_	_	_	_	_	_	_	_	_	_	_	_	1
H16	_	_	1	_	_	_	_	_	_	_	_	_	_	_	_	_	1
H17	_	_	1	_	_	_	_	_	_	_	_	_	_	_	_	_	1
H18	_	_	_	_	_	_	_	_	1	_	_	_	_	_	_	_	1
H19	_	_	_	_	_	_	_	_	1	_	_	_	_	_	_	_	1
H20	_	_	_	_	_	_	_	_	1	_	_	_	_	_	_	_	1
H21	_	_	_	1	_	_	_	_	_	_	_	_	_	_	_	_	1
H22	_	_	_	1	_	_	_	_	_	_	_	_	_	_	_	_	1
H23	_	_	_	1	_	_	_	_	_	_	_	_	_	_	_	_	1
H24	_	_	_	1	_	_	_	_	_	_	_	_	_	_	_	_	1
H25	_	_	_	1	_	_	_	_	_	_	_	_	_	_	_	_	1
H26	_	_	_	_	_	_	_	_	_	_	_	_	_	_	3	_	3
H27	_	_	_	_	_	_	_	_	_	_	_	_	_	_	1	_	1
H28	_	_	_	_	_	_	_	_	_	_	_	_	_	_	1	_	1
H29	_	_	_	_	_	_	_	_	_	_	_	_	_	_	1	1	2
H30	_	_	_	_	_	_	_	_	_	_	_	_	_	1	_	_	1
H31	_	_	_	_	_	_	_	_	_	_	_	_	_	1	_	_	1
H32	_	_	_	_	_	_	_	_	_	_	_	_	_	2	_	_	2
H33	_	_	_	_	_	_	_	_	_	_	_	_	1	_	_	_	1
H34	_	_	_	_	_	_	_	_	_	1		1	1	_	_	_	3
H35	_	_	_	_	_	_	_	_	_	_	_	_	1	_	_	_	1
H36	_	_	_	_	_	_	_	_	_	_	_	_	1	_	_	_	1
H37	_	_	_	_	_	_	_	_	_	1	_	_	_	_	_	_	1
H38	_	_	_	_	_	_	_	_	_	1	_	_	_	_	_	_	1
H39	_	_	_	_	_	_	_	_	_	1	_	_	_	_	_	_	1
H40	_	_	_	_	_	_	_	_	_	1	_	_	_	_	_	_	1

A. Haplotypes from EC lineage, continued

	SN	BU	KA	KO	GF	WK	EK	GB	KK	MP	ZO	MT	MM	MLZ	KS	NK	Total
H41	_	_	_	_	_	_	_	_	_	1	_	_	_	_	_	_	1
H42	_	_	_	_	_	_	_	_	_	1	_	_	_	_	_	_	1
H43	_	_	_	_	_	_	_	_	_	_	_	1	_	_	_	_	1
H44	_	_	_	_	_	_	_	_	_	_	_	1	_	_	_	_	1
H45	_	_	_	_	_	_	_	_	_	_	_	1	_	_	_	_	1
H46	_	_	_	_	_	_	_	_	_	_	_	2	_	_	_	_	2
H47	_	_	_	_	_	_	_	_	_	_	_	1	_	_	_	_	1
H48	_	_	_	_	_	_	_	_	_	_	_	1	_	_	_	_	1
H49	_	_	_	_	_	_	_	_	_	_	_	1	_	_	_	_	1
H50	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	3	3
H51	3	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	3
H52	1	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	1
H53	1	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	1
H54	_	_	_	_	_	_	_	_	_	_	1	_	_	_	_	_	1
H55	_	_	_	_	_	_	_	_	_	_	1	_	_	_	_	_	1
H56	_	_	_	_	_	_	_	_	_	_	1	_	_	_	_	_	1

Table 3.3 continued**. B.** Haplotypes from KZN lineage

	KK	UZ	MD	UM	MZ	SH	MQ	Total
H1	_	_	1	_	_	_	-	1
H2	_	_	1	_	_	_	-	1
H3	_	_	2	_	_	_	-	2
H4	_	_	1	_	_	_	-	1
H5	_	_	_	1	1	_	-	2
H6	1	_	_	1	1	_	-	3
H7	_	_	_	_	_	6	-	6
H8	_	_	_	_	_	2	-	2
H9	_	1	_	_	_	_	-	1
H10	_	1	_	_	_	_	-	1
H11*	3	1	_	5	_	_	-	9
H12	_	1	_	_	_	_	-	1
H13	2	1	_	1	_	_	-	4
H14	_	1	_	1	_	_	-	2
H15	_	1	_	1	_	_	-	2
H16	1	_	_	_	_	_	-	1
H17	1	_	_	_	_	_	-	1
H18	-	-	-	-	-	-	1	1

Table 3.4. Comparison of pairwise Φ_{ST} distance values between all the localities in EC lineage based on cytochrome *b* sequences. The star marked localities yielded the comparisons that were significantly high ($P < 0.05$). Locality abbreviations follow Table 3.1.

	SN	**BU***	**KA**	**KO**	**GF***	**WK**	**EK**	**GB**	**KK***	**MP**	**ZO**	**MT**	**MM**	**MLZ**	**KS**
SN	-														
BU	0.05	-													
KA	0.17	0.19	-												
KO	0.07	0.15	0.03	-											
GF*	0.53	0.55	0.32	0.36	-										
WK	0.23	0.33	0.13	0.02	0.83	-									
EK	0.13	0.23	0.03	-0.06	0.46	0.03	-								
GB	0.19	0.28	0.08	0.01	0.41	0.16	0.02	-							
KK	0.16	0.27	0.13	0.09	0.11	0.27	0.07	0.04	-						
MP	0.22	0.28	0.16	0.09	0.37	0.19	0.08	0.05	0.09	-					
ZO	0.21	0.27	0.07	0.001	0.79	0.13	0.01	0.22	0.29	0.17	-				
MT	0.21	0.31	0.18	0.08	0.62	0.01	0.04	0.16	0.28	0.18	0.05	-			
MM*	0.36	0.40	0.36	0.33	0.61	0.39	0.30	0.33	0.37	0.31	0.27	0.35	-		
MLZ	0.19	0.28	0.16	0.07	0.67	0.08	0.05	0.16	0.26	0.16	0.06	0.05	0.28	-	
KS*	0.39	0.43	0.39	0.35	0.67	0.44	0.35	0.38	0.42	0.36	0.34	0.39	0.37	0.18	-
NK*	0.44	0.46	0.39	0.37	0.83	0.60	0.41	0.49	0.49	0.39	0.46	0.46	0.37	0.25	0.15

Table 3.5. The source of variation, degrees of freedom, sum of squares, variance components and percentage of variation among different groups, among localities within groups and within localities for the EC lineage that resulted from the AMOVA analysis of cytochrome *b* data.

Source of variation	**Degrees of freedom**	**Sum of squares**	**Variance components**	**Percentage variation**
Among groups	2	41.49	1.27	38.66
Among localities within groups	13	57.54	0.32	9.76
Within localities	115	194.60	1.69	51.58
Total	130	293.63	3.28	

Table 3.6. Comparison of pairwise Φ_{ST} distance values between all the localities in KZN lineage based on cytochrome *b* data. The star marked localities yielded the comparisons that were significantly high ($P < 0.05$). Locality abbreviations as in Table 3.1.

	KK	**UZ**	**MD***	**UM**	**MZ**	**MQ**
KK	-					
UZ	0.07	-				
MD*	0.41	0.37	-			
UM	-0.01	0.08	0.43	-		
MZ	-0.08	-0.06	0.34	-0.11	-	
SH*	0.67	0.59	0.77	0.62	0.75	
MQ*	0.95	0.93	0.96	0.96	0.96	-

Table 3.7. The source of variation, degrees of freedom, sum of squares, variance components and percentage of variation among different groups, among localities within groups and within localities for the KZN lineage that resulted from the AMOVA analysis of cytochrome *b* data.

Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage variation
Among groups	2	68.7	3.20	73.7
Among localities within groups	4	12.1	0.41	9.37
Within localities	32	23.5	0.74	16
Total	38	104.4	4.35	93

Table 3.8. List of localities, the lineage in which each occurs, the number of sequences and haplotypes per locality as well as the haplotype and nucleotide diversity of the S7 intron 1 marker for each locality (In determining *h* and π diversities gaps in the sequences were treated as a fifth state). Locality abbreviations are as assigned in Table 3.1.

Locality	Lineage	Number of sequences	Number of haplotypes	Haplotype diversity (*h*)	Nucleotide diversity (π)
SN	EC	3	2	0.66	0.0015
KA	EC	10	4	0.73	0.0015
KO	EC	5	3	0.70	0.0009
GF	EC	9	2	0.39	0.0005
EK	EC	5	4	0.90	0.0039
KK	EC and KZN	12	11	0.98	0.0098
UZ	KZN	2	1	0.00	0.0000
MP	EC	7	3	0.52	0.0007
ZO	EC	7	6	0.95	0.0091
MT	EC	6	5	0.93	0.0029
MM	EC	3	3	1.00	0.0023
MD	KZN	4	2	0.67	0.0008
UM	KZN	8	8	1.00	0.0045
MLZ	EC	6	6	1.00	0.0030
SH	KZN	5	4	0.90	0.0049
KS	EC	4	2	0.67	0.0008
NK	EC	3	3	1.00	0.0047
MQ	KZN	3	2	0.67	0.0008

Table 3.9. Distribution of haplotype diversity from the nuclear S7 intron 1 sequences of *Glossogobius callidus* specimens among localities. Population abbreviations are according to Table 3.1. Haplotypes with the highest number of representatives and possibility of being ancestral are marked with a *.

	SN	KA	KO	GF	EK	KK	UZ	MP	MT	ZO	MD	UM	MM	MLZ	KS	NK	SH	MQ	Total
H1*	-	4	3	7	2	-	-	-	-	1	-	-	-	-	-	-	-	-	17
H2*	1	1	1	2	1	-	-	-	-	2	-	-	-	-	-	1	-	-	9
H3	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1
H4	-	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4
H5	-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	2
H6	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
H7	-	-	-	-	-	1	-	-	-	-	--	-	-	-	-	-	-	-	1
H8	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-		-	1
H9	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
H10	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
H11	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
H12	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
H13	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
H14	-		-	-	-	1		-	-	-	-	-	-	-	-		-		1
H15-	-	-	-	-	-	2	-	-	-	1	-	-	-	-	-	-	-	-	3
H16	-	-	1	-	-	-	-	-	-	-		-	-		-	-	-	-	1
H17	-	-	-	-	-	-	-	-	-	-		-	-	1	2	-	-	-	3
H18	-	-	-	-	-	-		-	-	-		-	-	1	2	-	-	-	3
H19	-	-	-	-	-	-	1	-	-	1	2	1	-	-	-	-	-		5
H20*	-	-	-	-	-	-	-	-	-	-	2	1	-	-	-	-	-	2	5
H21*	-	-	-	-	-	-	-	5	2	-	-	-	-	1	-	-	-	-	8
H22	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	1
H23	-	-	-	-	-	-	-		--	-	-	-	-	1	-	-	-	-	1
H24	-	-	-	-	-	-	-	-	1	-	-	-	-	1	-	-	-	-	2
H25	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	1
H26	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	1
H27	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	1
H28	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	1

Table 3.9. continued

	SN	KA	KO	GF	EK	KK	UZ	MP	MT	ZO	MD	UM	MM	MLZ	KS	NK	SH	MQ	Total
H29	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	1
H30	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1
H31	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1
H32	-	-	-	-	-	-	-		1	-	-	-	-	-	-	-	-	-	1
H33	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1
H34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1
H35	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	2
H36	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1
H37	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1
H38	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1
H39	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	2
H40	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	1
H41	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	1
H42	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	1
H43	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	1
H44	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	1
H45	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	1
H46	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1
H47	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1
H48	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1
H49	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1

Table 3.10. Comparison of pairwise Φ_{ST} distance values between all the analysed localities for the S7 intron 1 sequences. The star marked localities yielded comparisons that were significantly high ($P < 0.05$). Locality abbreviations are as assigned in Table 3.1

	SN	**KA**	**KO**	**GF**	**EK**	**MP***	**MT***	**MM***	**MLZ**	**KS***	**NK***	**ZO**	**UZ***	**MD***	**UM***	**SH***	**MQ***	**KK***
SN	-																	
KA	0.22	-																
KO	0.19	-0.03	-															
GF	0.36	0.15	-0.09	-														
EK	0.002	0.08	-0.03	0.06	-													
MP*	0.63	0.54	0.59	0.67	0.39	-												
MT*	0.49	0.51	0.50	0.61	0.32	0.38	-											
MM*	0.52	0.56	0.58	0.71	0.32	0.56	0.45	-										
Mlz	0.12	0.20	0.14	0.25	0.02	0.33	0.19	0.36	-									
KS*	0.62	0.57	0.61	0.72	0.26	0.78	0.50	0.71	0.13	-								
NK	0.08	0.17	0.14	0.31	-0.01	0.51	0.36	0.34	0.10	0.46	-							
ZO	0.14	0.35	0.27	0.38	0.22	0.43	0.38	0.32	0.26	0.34	0.10	-						
UZ*	0.93	0.92	0.96	0.98	0.82	0.97	0.87	0.92	0.86	0.97	0.80	0.27	-					
MD*	0.93	0.92	0.95	0.97	0.85	0.96	0.88	0.92	0.87	0.96	0.85	0.35	0.11	-				
UM*	0.76	0.83	0.81	0.86	0.76	0.85	0.79	0.79	0.78	0.81	0.72	0.39	0.03	0.12	-			
SH*	0.78	0.85	0.83	0.88	0.76	0.87	0.80	0.80	0.79	0.83	0.73	0.35	0.18	0.17	0.01	-		
MQ*	0.92	0.91	0.94	0.97	0.83	0.96	0.87	0.91	0.86	0.96	0.82	0.30	0.68	0.20	0.15	0.07	-	
KK*	0.33	0.50	0.42	0.50	0.41	0.51	0.51	0.45	0.44	0.46	0.33	0.03	0.11	0.18	0.26	0.19	0.14	0

Table 3.11. The source of variation, degrees of freedom, sum of squares, variance components and percentage of variation among different groups, among localities within groups and within localities for the S7 nuclear intron 1 marker that resulted from the AMOVA analysis.

Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage variation
Among groups	1	193.53	4.14	65.63
Among localities within groups	16	76.11	0.58	9.15
Within localities	84	133.68	1.59	25.22
Total	101	403.32	6.31	

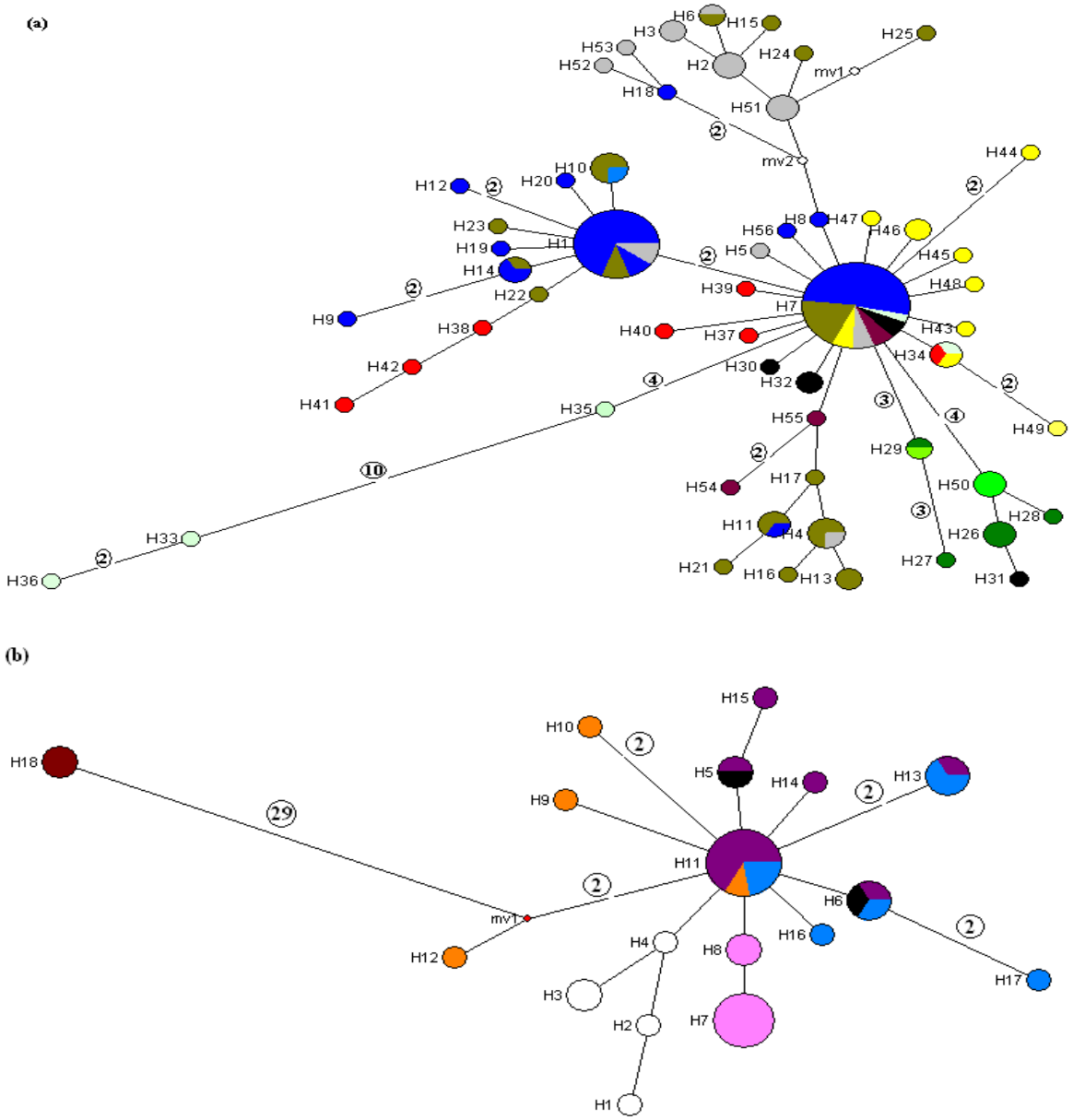


Figure 3.1. The haplotype network of *Glossogobius callidus* cyt *b* sequences (a) EC lineage showing four geographic groups and a star-like phylogeny in association with H1 and H7, (b) KZN lineage showing only a star-like phylogeny in association with H11. Each circle indicates a unique haplotype while the colours represent the different localities ((a) Warm-temperate =: ■ Bushmans + Sundays, ■ Kariega + Kowie, ■ (Great Fish + Keisekamma + Gonubie + East and West Kleinemonde); subtropical =: ■ Mtentweni, ■ Lake Mzingazi, ■ Mpenjati ■ Zotcha, ■ Nkanini + Kosi Bay, ■ Mzimayi,: (b) Tropical: ■ Mozambique,Subtropical =: ■ Mzumvubu, ■ Shengeze Lake, □(white) Mdloti, ■ Mhlatuzi, ■ Mzingazi canal, Warm-temperate: ■ Keisekamma only). Numbers of mutational differences are only displayed for haplotype links with two or more mutations.

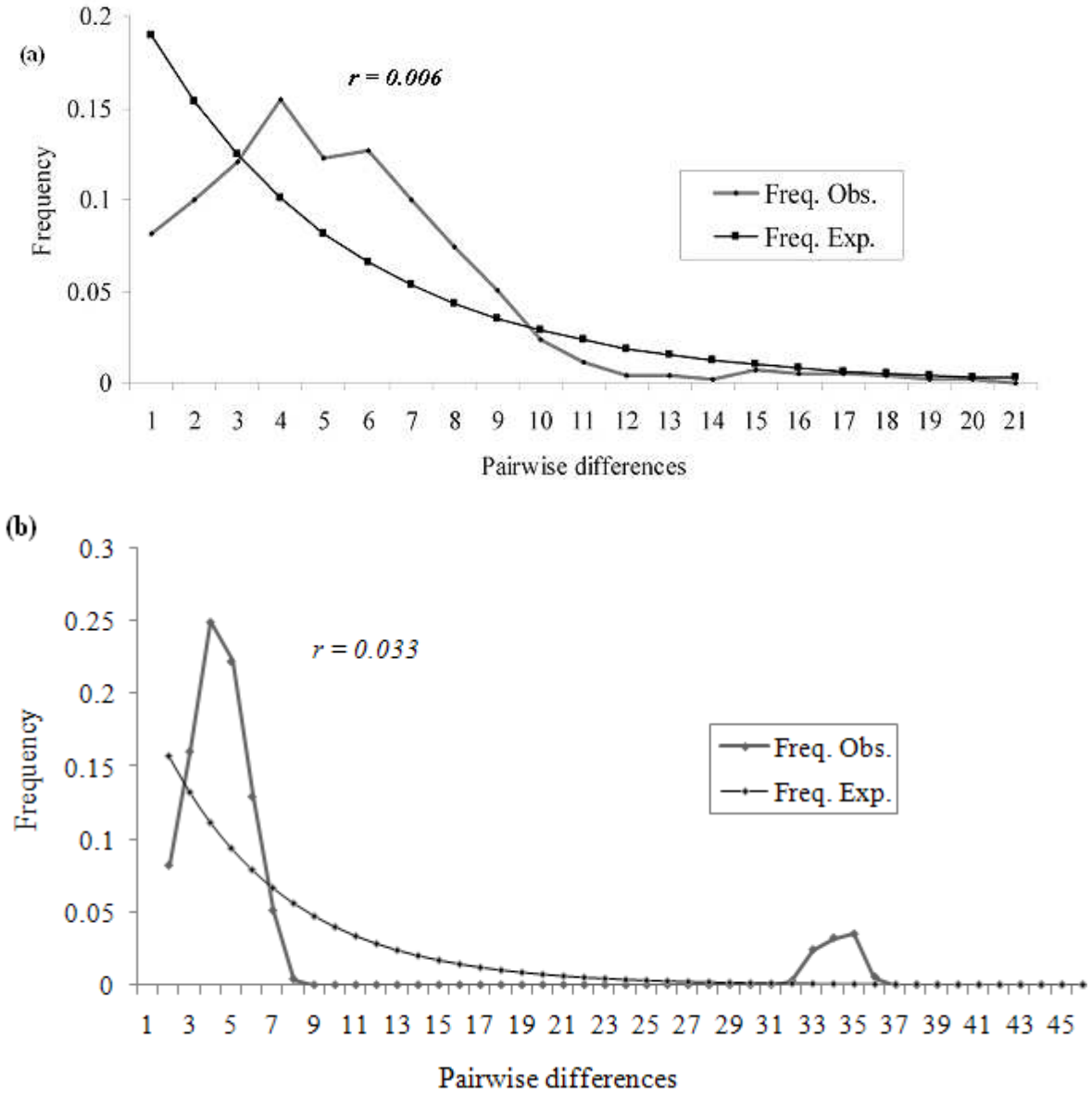


Figure 3.2. The unimodal mismatch distribution for cytochrome *b* sequences. a) EC lineage, mean pairwise difference = 4.28, and b) KZN lineage, mean pairwise difference = 4.33. The Raggedness statistics (*r*) observed with the unimodal distribution in EC lineage (a) and bimodal KZN lineage (b) as well as the observed and expected frequency distributions are indicated.

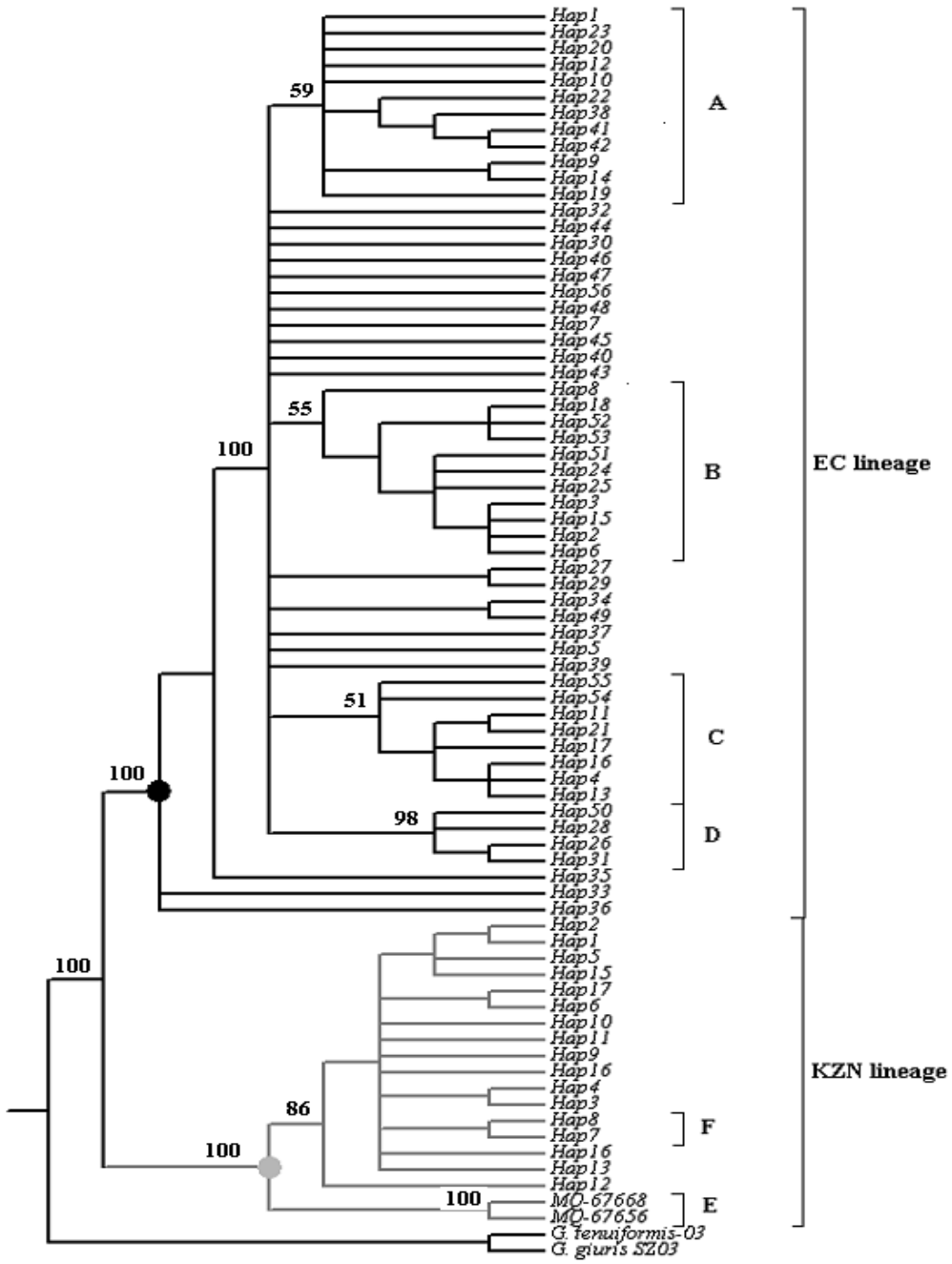


Figure 3.3. Maximum-likelihood estimate of the phylogenetic relationship among the *Glossogobius callidus* populations included in the study using the TrN+G model (-ln likelihood = 3776.07; shape parameter b = 0.19). Bold numbers at nodes are bootstrap support for MP searches at 1000 bootstrap replicates. Only nodes with a support >50% are shown. The tree indicates two distinct lineages within *G. callidus*. A = Great Fish/Kleinemond/Gonubie/Keisekamma, B = Bushmans/Sundays, C = Kariega/Kowie, D = Kosi Bay/Nkanini, E = Mozambique, and F = Shengeze lineages.

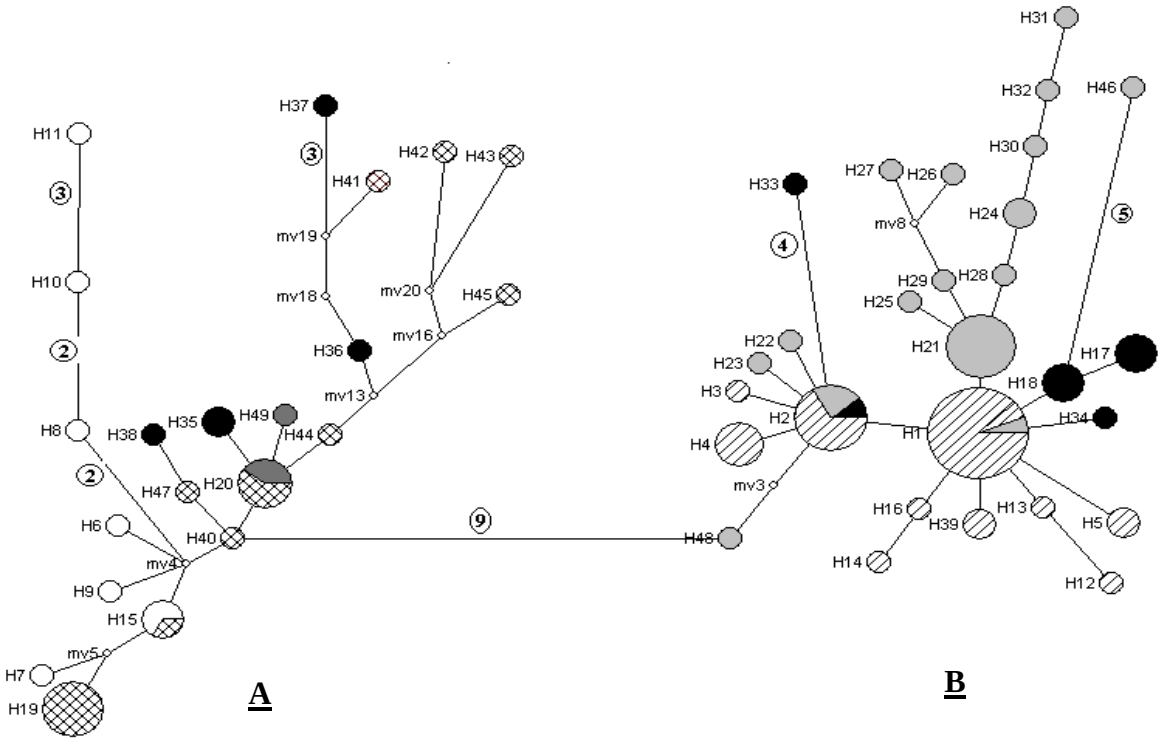


Figure 3.4. The haplotype network of *Glossogobius callidus* S7 nuclear intron 1 sequences showing two lineages (KZN (A) and EC (B) lineages) of the tested individuals. Each circle indicates a unique haplotype while the colours represent the different geographic structuring [■ KZN localities within EC (B) lineage (Mpenjati, Mtentweni, Zotcha and Lake Mzingazi), ■ Nkanini/ Kosi bay/ Shengeze (localities -10km apart), (white) Keisekamma (EC locality within KZN lineage), ■ Mozambique (dark grey), (backward stripes) Warm-temperate, Eastern Cape localities (Sundays, Keisekamma, Kowie, Great Fish, East Kleinemonnde and Keisekamma), and (cross checked) Subtropical, KZN localities (Mhlatuzi, Mzumvubu, Mdloti and Mzimayi]. Numbers of mutational differences are only displayed in a circle for haplotypes linked with two or more mutations.

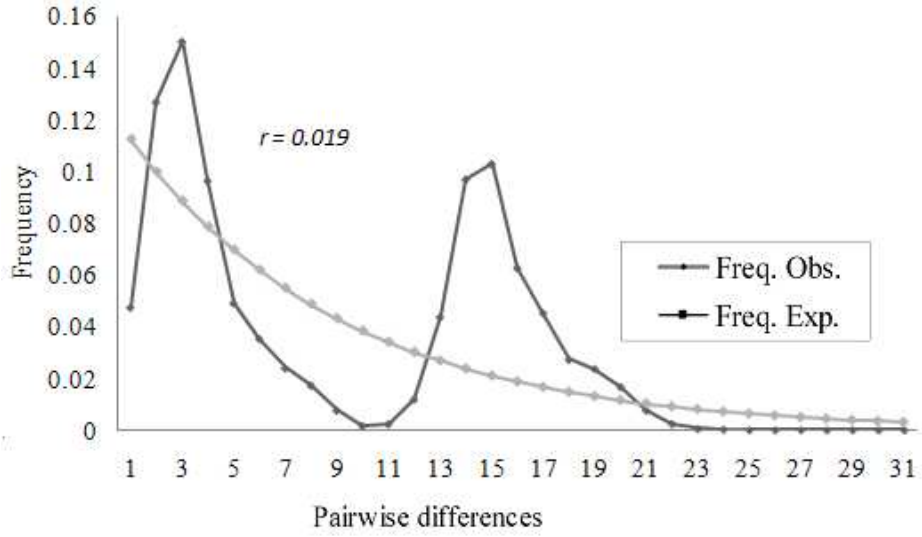


Figure 3.5. The bimodal mismatch distribution for S7 nuclear intron 1 sequences. The mean pairwise difference between two peaks = 7.89. The Raggedness statistic (*r*) observed with the mismatch distribution and the observed and expected frequency distributions are shown.

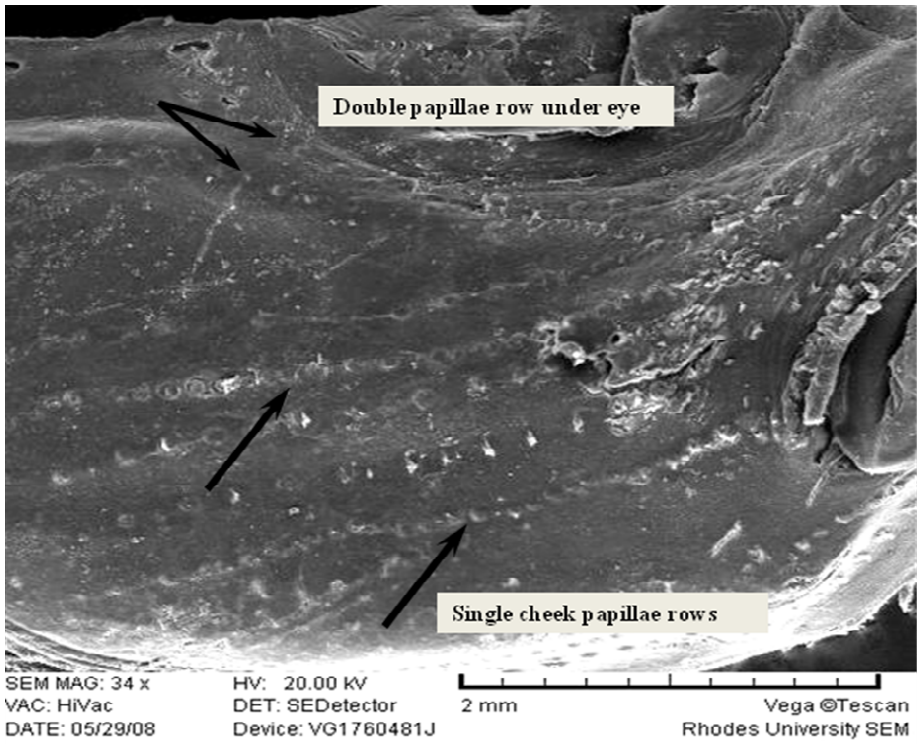


Figure 3.6. Cheek papillae row patterns of *Glossogobius callidus* Smith, 1937 (specimen number: GCKS06) obtained with the aid of scanning electron microscopy. Note the doubled papillae row under the eye, and the single cheek papillae rows.

CHAPTER 4

PHYLOGENETIC RELATIONSHIPS OF *GLOSSOGOBIUS* SPECIES OCCURING IN THE SOUTHERN AFRICAN REGION

4. INTRODUCTION

The genus *Glossogobius* was first erected by Gill, 1862 and currently contains 22 species (Nelson 2006), which are distributed throughout the tropical Indo-west Pacific in freshwater streams, rivers and brackish estuaries (Pusey *et al*. 2004). Most species of the genus in southern Africa are found in estuaries, and in the lower and upper reaches of the rivers (Skelton 2001; Pusey *et al*. 2004). These habitats are characterised by a high degree of instability and are influenced by terrestrial events such as rainfall and soil erosion, and marine events such as currents, tides and waves (Hoese 1986; Goren 1996; Whitfield 1998). The review of the genus by Hoese and Winterbottom (1979) followed by Hoese (1986) reported three species of *Glossogobius* (*G. callidus, G. giuris* and *G. biocellatus*) in South Africa. Of the three, *Glossogobius giuris* and *Glossogobius callidus* are the most widely distributed and common species of this genus. The "third" species of the genus has subsequently been moved to another genus, namely *Psammogobius* (Bauchot *et al.* 1991). New morphological work done on the southern African *Glossogobius* species (Helen Larson; Art Gallery and Museum of the Northern Territory (*pers. comm*)) indicated that there are still three species of the genus in southern Africa (*G. callidus, G. giuris* and *G. tenuiformis*). The third "nominal" species of Fowler (1934), *G. tenuiformis* was originally described from Lake St Lucia, KwaZulu-Natal. Following the work of Smith (1960) and Hoese and Winterbottom (1979), Hoese (1986) had reduced *G. tenuiformis* to a junior synonym of *G. giuris.* Therefore, *G. giuris* and *G. callidus* are currently the only two valid species of this genus in southern Africa. Unlike the southern African endemic *G. callidus, G. giuris* is also widely distributed throughout the Indo-west Pacific (Skelton 2001; Pusey *et al*. 2004) where it has a patchy distribution in rivers of wet tropical regions, never reaching high abundance in these systems (Pusey *et al*. 2004).

Although, several systematic studies on the phylogeny and taxonomy of gobies have been undertaken, these have been generally based on a variety of morphological characters (Harrison 1989; Hoese and Gill 1993). It has however been noted by various scientists (e.g.

Akihito and Meguro 1975; Pusey *et al*. 2004) that the diagnostic morphological characters that are commonly used for species delineation vary within species distributions (Ridgway *et al.* 1998; Thacker 2003). It has therefore been difficult or impossible for systematists to accurately discern species boundaries for gobiid groups because of this intraspecific variation (Akihito *et al.* 2000; Thacker and Cole 2002). This problem of variation within a single species has been observed within the genus *Glossogobius* and thus resulted in synonymies of its species (Hoese and Winterbottom 1979; Hoese 1986). Because of the difficulty in finding diagnostic morphological features in these gobies, it has been generally considered that synonymies such as *G. tenuiformis* may represent valid species. Cryptic species are usually prevalent and have been identified in widespread species (Indo-Pacific in this case) that exhibit morphological variation. It is therefore possible that *G. giuris* is a species complex in need of revision (see Akihito and Meguro 1975). The treatment of variation in certain morphological characters (e.g. counts of pectoral fin rays and predorsal scales, and cheek papillae row patterns) could have caused other species such as *G. tenuiformis* to be synonymised with *G. giuris* (see Akihito and Meguro 1975). The taxonomic status of *G. tenuiformis* is reviewed in this study using specimens collected in KwaZulu-Natal following the original description for this synonymy. This analysis is also supported by earlier results (Chapter 3) on the genetic structure of the morphologically variable *G. callidus* that have indicated two unique lineages that could also represent two species.

Before the advent of molecular systematics, morphological characters were generally used for inferring fish phylogenetic relationships. Molecular characters however have since been very useful in providing a rich source of information for analysis of species phylogeny. This is especially more evident in cases where it has been difficult to differentiate species because of the similarities in external morphology and diagnostic characters as in the case of gobies (Hoese 1986; Thacker 2003). Molecular analyses can also within limits, indicate the approximate timing of species separation as well as the sequence of evolutionary events (Hillis and Moritz 1990). Several recent studies (Akihito *et al*. 2000; Harada *et al*. 2002; Lavoue *et al*. 2003; Carlsson *et al*. 2004; Coronel *et al*. 2004; Morrison *et al*. 2006; Kuriiwa *et al*. 2007; Haponki and Stepien 2008) have demonstrated that combining mtDNA and nuclear gene data can produce better and more robust phylogenetic resolution between closely related taxa. Additionally, Gadagkar *et al*. (2005) stated that it is the concordance of phylogenetic patterns across several markers that provide the most insightful views of phylogenies and their history of divergence. Therefore, nucleotide sequence data (cytochrome

b and S7 ribosomal protein gene intron 1) were used as a character source in the present study to investigate the taxonomic status of *G. tenuiformis* and to hypothesise about the relationships between the species of the genus in southern Africa.

4.1. Materials and methods

All materials and methods used are as described in Chapter 2 under materials and methods.

4.2. Results

4.2.1 Cytochrome *b*

Of the 1045 nucleotide sites examined, 612 characters were constant, 36 characters were parsimony-uninformative (3.4 %), and 397 characters were parsimony-informative (37.9%). Heuristic search analysis of parsimony (MP) produced 28 most parsimonious trees with the tree length of 1271 steps (Figure 4.1), calculated consistency index (CI = 0.54), retention index (RI = 0.89), and re-scaled consistency index (RC = 0.48). The moderate high values of the RI imply that there are minimal homoplasy characters in the data sets and thatthe estimated cladogram retains a strong phylogenetic signal. The monophyly of the genus is strongly supported by 90% bootstrap support. Bootstrap analysis also indicated strong support for the basal nodes that define the six major clades and relationships among clades, while terminal nodes receive relatively moderate support with a few exceptions. Inter-clade sequence divergences ranged from 3% between Australian *G. giuris* and African *G. giuris* (Figure 4.1, clade D and E respectively), reaching up to 10% sequence divergence between *G. callidus* (clade A and B) and *G. tenuiformis* (clade C) (Figure 4.1). An Australian DNA specimen labelled as *G. aureus01* grouped with the Australian *G. giuris* group (E) (Figure 4.1) could be a misidentification that requires verification and will therefore be treated as such in further results. All of the species nodes have bootstrap support of a 100% suggesting that these recognised species are genetically distinct. However, some internal nodes have moderate to strong bootstrap support ranging from 51 to 97% which may suggests some genetic structuring or cryptic species within these species. For example, within the KZN lineage, the South African population of *G. callidus* was about 1.2 % diverged from the haplotypes representing Mozambique population (Mq. 67668 and Mq. 67656.4) (for more details refer to Chapter 3).

Maximum likelihood and Bayesian inference analyses were also performed on these datasets. ML was based on a model selected under AIC in Modeltest (TrN+G, gamma correction = 0.166, and nucleotide composition as follows: A = 26%, C = 35%, G = 14%, and T = 25%), and recovered a single tree (Figure 4.2) (-ln*L* = 6744.5). Due to the time constrains in bootstrapping the likelihood tree, the ML analyses were not bootstrapped. There was congruence between Bayesian Inference (BI) (not shown) and the ML analyses (Figure 4.2), as they estimated the same basic topology. Additionally, the monophyly of all the species is highly supported (≥ 100% posterior probabilities) by the Bayesian Inferrence. In general, the three phylogenetic analyses all recover four well-supported clades (Figure 4.1 and 4.2: clades C, F, G and H), corresponding to the different recognised species including the synonymy *G. tenuiformis*. Although the trees include *Glossogobius* species from Australia, there is no obvious separation or geographic origin of southern African taxa as the individuals from these two areas share a recent common ancestor (e.g. clade D and E) (Figure 4.1 and 4.2). However, the inclusion of other taxa of the genus could assist in resolving the placement of these groups. Clade A, which is well supported by all analyses, comprised of the widely distributed *G. callidus* population collected from both the subtropical and warm-temperate regions of South Africa, including the type locality of the species. Clade B is a monophyletic assemblage comprising of *G. callidus* individuals from the subtropical region of South Africa (defined as Eastern Cape lineage in chapter 3), reaching far into the streams and rivers of Mozambique. Apart from the reported 10% sequence divergence between *G. callidus* and *G. tenuiformis*, high sequence divergence (8.5% divergence) is also observed between the "Australian *G. giuris* population" (Figure 4.1, group E) and African *G. callidus* (Clade B). All the monophyletic clades (A to H, with the exception of group E in the MP topology) are well supported by the posterior probability (100%) and MP bootstrap values (>90%). In summary the cytochrome *b* analyses of the dataset supported the monophyly of the genus *Glossogobius* (90% bootstrap and 100% posterior probability) and the recognition of each major clade to a species level (Figures 4.1 and 4.2).

However, the ML and BI analyses did recover some species relationships that were not evident in MP. For example, these analyses (ML and BI, see Figure 4.2) suggested that *G. giurus* (clade G) is more basal in the genus to the other analysed taxa. The ML analysis also suggested (even not well supported by the posterior probability) that the southern African *G. callidus* complex (clade A and B) is the sister group of the undescribed *Glossogobius* species (clade F) collected from the Daly River estuary in the Northern Territory, Australia (Figure

4.2). There is 5.8% and 7.2% sequence divergence between this Australian species (clade F) and the two *G. callidus* clades (clades A and B) respectively (Figure 4.1 and 4.2). All three clades (A, B and F), probably representing three species possibly share a most recent common ancestor.

It appears that each southern African *Glossogobius* species has a sister species occurring in Australia. In both BI and the ML analysis *G. aureus* (*G. aureus02*) was recovered as the sister species of the South African *G. tenuiformis* (clade C) (100% posterior probability, Figure 4.2). The total DNA sequence divergence between the two species was also at its lowest at 3.0%. Surprisingly, the divergence between African *G. giuris* (clade D) and the Australian *G. giuris* (group E) was also the lowest at 3.0% sequence divergence and may be the reason for the low support for group E in figure 4.1. According to ML and BI analyses, both populations of *G. giuris* are monophyletic, with a clear separation between the Australian and the African *G. giuris* lineages (Figure and 4.2). The genetic divergence between the African *G. giuris* (clade D) and the other southern African species of *G. callidus* (clade H) and *G. tenuiformis* (clade C) is 6.5% and 6.9%, respectively, which is higher than the 3% between the African and Australian populations of *G. giuris*.

4.2.2 S7 ribosomal protein gene, intron 1

For the S7 intron 1, only 41 individuals representing the three *Glossogobius* species found in southern Africa were included in the analyses. Analyses of sequences with gaps treated as present /absent data and the exclusion of gaps from the analyses produced the same topology with no difference in the nodal support for the species and subdivisions. Therefore, only results with all gaps removed from the data are presented for this marker. Due to difficulties in amplifying the S7 intron marker, the specimens collected from Australia were not successfully amplified and thus not included in the analyses. Additionally not all individuals used in the analyses of cytochrome *b* could successfully be amplified and therefore not all are present in this analysis either. Of the 668 available characters, 91 were constant (13.6%), 122 characters were parsimony-uninformative (18.3%) and 455 were parsimony-informative (68.1%).

Heuristic search analyses of MP (Figure 4.3) produced 12 most parsimonious trees with a tree length of 887 steps. The calculated CI = 0.90, RI = 0.97, and re-scaled consistency index (RC) = 0.87 are higher than in the cytochrome *b* analyses indicating a much lower frequency

of homoplasious characters and a better retention of the phylogenetic signal. Bootstrap analysis only revealed strong support among the basal nodes that defined the major clades and relationships (clades A to D, and H), while terminal intraspecific nodes were not well supported. Using the sequence of *Bathygobius curacao* as outgroup, the southern African *Glossogobius* species are divided into two highly supported (100% bootstrap) major groups, the first (clade H) with both *G. callidus* lineages (clade A and Clade B), and the second consisting of *G. tenuiformis* (clade C) and *G. giuris* (clade D). Basal to these groups (clades C, D and H) were the two sequences of *Glossogobius celebuis* supported by 100% bootstrap (Figure 4.3). As already revealed by the cytochrome *b* trees, *G. callidus* was monophyletic (100% bootstrap) but separated into two lineages with 80 and 63 % bootstrap support values respectively (clades A and the polytomy B).

Modeltest identified K81+G (-lnL = 3833.1, gamma = 7.034) as the model that best fits the data under AIC. Using the identified model of evolution the recovered ML tree (-lnL = 3817.52, not shown) was almost identical to the BI topology. The results of the MP and BI (Figure 4.4) tree analyses are also similar. All analyses recovered the monophyly of *G. tenuiformis* (clade C, 100% posterior probability and MP bootstrap) and *G. giuris* (Clade D, 100% posterior probability and MP bootstrap). The monophyly of clade A (*G. callidus* sp. 1) was less supported (MP 80% bootstrap, Figure 4.3 and BI 91% posterior probability, Figure 4.4) than with the cyt *b* results. However, the monophyly of individuals of clade B (*G. callidus* sp. 2) was not well supported in MP (63% bootstrap, Figure 4.3) and not recovered in BI and ML analyses, with the individuals forming unresolved polytomies, (B, Figure 4.4).

4.2.3 Combined data (Cytochrome *b* and S7 intron1)

For the combined dataset no outgroups were available, therefore only the southern African specimens were used in this dataset. Midpoint rooting was applied to all analyses (MP, BI and ML). The combined dataset with gaps was 2047 base pairs long. When gaps were removed from the dataset, only 1659 characters were present. Analyses of datasets with or without gaps revealed similar tree topologies and branch support of the clades and relationships between them hence only trees computed using the dataset without indels are presented.

The heuristic search of MP (1659 base pairs) identified 1222 characters as constant, 19 variable characters being parsimony-uninformative, and 418 characters parsimony-

informative. Fifty one trees were recovered with a length of 609 (CI = 0.84, RI = 0.98, and RC = 0.82). The ML tree analysis was computed using TrN+I+G (gamma = 0.71, base frequencies: A = 22%, C = 29%, G = 23%, and T = 26%) as the model of evolution chosen under AIC in Modeltest. The Bayesian tree topology was the same as that estimated with ML analysis. Since all three analyses produced similar tree topologies, only the ML topology is presented (Figure 4.5) with parsimony bootstrap percentages and posterior probability values.

Although midpoint rooting was applied to the combined dataset (cytochrome *b* and S7 intron1), the analyses recovered four clades (clade A = *G. callidus* sp.1 (100% bootstrap and posterior probability), clade B = *G. callidus* sp.2 (100% bootstrap and 99% posterior probability), clade C = *G. tenuiformis* (100% bootstrap and posterior probability) and clade D = *G. giuris* (100% bootstrap and posterior probability)) (Figure 4.5) which correspond well with the three described species and possibly a new species of the flat head gobies found in southern Africa region. The monophyly of the two previously recognised lineages of *G. callidus* (clade H) is recovered and strongly supported (100% bootstrap and posterior probability) in all analyses. Additionally, the monophyly of *G. callidus* sp.1 (clade A, 100% bootstrap and posterior probability) and *G. callidus* sp.2 (100% bootstrap and 99% posterior probability) were also well supported. As in the S7 intron 1 marker, the monophyly of clade C and D (*G. tenuiformis* and *G. giuris*, respectively) is also well recovered in all three analyses (100% bootstrap and posterior probabilities) (Figure 4.5).

4.2.4 Scanning Electron Microscopy (SEM)

As previously mentioned the configuration of the series of sensory papillae on the cheeks of gobies has been used in the identification of genera and species of this family (Akihito and Meguro 1975). The *G. giuris* specimen had two doubled cheek papillae rows and a single row of papillae under the eye (Figure 4.6, top). *Glossogobius tenuiformis* also have the single row of papillae under the eye while the cheek papillae rows were all single (Figure 4.6, bottom). Individuals of the two genetically identified lineages of *G. callidus* were also prepared for SEM but probably due to the incorrect specimen preparations the cheek papillae pattern of most specimens were destroyed and the pappilae patterns could not be observed. However, obtained micrographs with visible cheek papillae (see Figure 3.6) could not be differentiated from each other as all had the same patterns of cheek papillae known for *G. callidus.*

4.3. Discussion and conclusions

The mitochondrial DNA (cyt *b*) dataset included more specimens than the S7 intron 1 due to successful amplification of the cyt *b* gene for most specimens. Specimens with data from only a single gene (cyt *b*) were excluded in the combined data analyses. Even though indels have been used successfully with promising results by various studies (Lavoue *et al*. 2003; Near *et al*. 2004; Kuriiwa *et al*. 2007), they were not considered as they did not add new information in the present study due to the similarities in the tree topologies and nodal supports when compared with the analyses using de-gapped data. However, the current results illustrate the phylogenetic effectiveness of the two classes of molecular markers at very different levels of relationship within the trees. For example, the S7 intron 1 provided strong phylogenetic information and resolution at the deeper clades among closely related species (Figure 4.3 and 4.4) whereas the mitochondrial genes provide abundant characters for phylogenetic inference on the intraspecific relationships within clades (Figure 4.1 and 4.2). The S7 intron 1 marker is evolving at a much slower rate than the cyt *b* and can thus be useful in solving deeper phylogenetic questions (e.g. at the genera and familial levels) and questions concerning the evolution of species (Simmons and Ocheterena 2000; Near *et al*. 2004). Most studies on intraspecific relationships have highlighted the importance of mtDNA in phylogenetic analyses (Harada *et al*. 2002; Haponki and Stepien 2008), and as such the mtDNA cyt *b* gene was used effectively in this study to resolve intraspecific relationships, especially *G. callidus*. As this study demonstrated, even though cyt *b* evolves at a much faster rate than the S7 intron 1 marker, it is also conserved enough to elucidate deeper and/or old relationships among species of the same genus, and thus its importance should not be ignored.

The molecular phylogenies estimated in this study clearly support the monophyly of the genus *Glossogobius* (Figure 4.1, 4.2, 4.3, and 4.4). This monophyly is strongly supported by the cyt *b* analyses with 90% bootstrap and 100% posterior probability values respectively (Figure 4.1 and 4.2). Although the MP analysis indicates that the analysed species of the genus were paraphyletic (Figure 4.1), there was no conflict with the ML/BI topology (Figure 4.2) with respect to the moderately to highly supported nodes (bootstrap value > 70% and BI posterior probabilities > 90%). The only minor difference between the MP and ML/BI topologies of the cyt *b* gene was regarding the monophyletic grouping of *G. aureus* with *G. tenuiformis*. In the ML/BI topology (Figure 4.2) *G. aureus* (*G. aureus02*) was the sister group of clade C (consisting of *G. tenuiformis*), while their sister group relationship was not

recovered in the MP tree (Figure 4.1). This uncertainty of the relationships could be attributed to the fact that *G. aureus* (here presented as *G. aureus02*) was represented by a single individual. Additionally, the sampling gap in this study, i.e. the absence of many other *Glossogobius* species, can also contribute to the observed results. Although in some studies incomplete lineage sorting was hypothesised as a possible cause of the observed paraphyly, unsatisfactory taxon sampling has been identified as the main cause of paraphyly in poorly known and undersampled genera or species (Funk and Omland 2003; Geurgas *et al*. 2008). Given the broad geographical distribution of *Glossogobius* species, more thorough sampling, including populations from areas not sampled in this study, will probably increase the resolution of relationships.

The southern African species are monophyletic and thus share a recent common ancestor based on the nuclear gene analyses with 100% bootstrap and posterior probability values (Figure 4.3 and 4.4). The results of the S7 intron 1 marker indicated that *G. celebius* (a freshwater and brackish species with an Asian and Oceania distribution), is basal to the southern African *Glossogobius* species; however this relationship within southern African taxa can only be confirmed with the addition of more taxa. The studied species of the genus in southern Africa form two genetically distinct and well-supported phylogroups, one represented by *G. callidus* populations, and the other comprising of *G. tenuiformis* and *G. giuris* (Figure 4.3 and 4.5). The two major clades can be characterised by some common morphological traits. Species of the basal clades C and D, representing *G. tenuiformis* and *G. giuris* respectively (Figure 4.4 and 4.5), have pre-dorsal scales reaching close behind the eye and few cycloid scales and a single row of papillae close under the eyes. The diagnostic cheek papillae patterns of these two species (i.e. cheek papillae rows double in *G. giuris* and all single in *G. tenuiformis*) were also observed on the micrographs of the SEM study (Figure 4.6). These characters were used by Hoese (1986) in his interpretations leading to different conclusions on the validity of *G. tenuiformis*, i.e. valid versus junior synonym of *G. giuris.* The species of clades A and B (*G. callidus* sp. 1 and sp. 2) are morphologically inseparable, all characterised by a naked pre-dorsal (i.e. no scales) area, distinct single or double blotch on caudal base and a double row of papillae directly under the eye (see Figure 3.6), while sharing characteristics such as their flat head, long snout and jaw and bilobed tongue with *G. tenuiformis* and *G. giuris*.

The cyt *b* tree topologies (Figure 4.2) positioned *G. giuris* (combined African and Australian populations, clade D and E) and *G. tenuiformis* (clade C) as distantly related lineages, while the S7 intron 1 (Figure 4.3 and 4.4) revealed that the two species (clade C and D) could be closely related. Both species have been genetically isolated as both their lineages are well-separated and strongly supported as unique monophyletic clades. In addition, the branch length separating the two species (Figure 4.5, combined data) as well as the cyt *b* percentage of sequence divergences between the two populations (6.5% divergence) also confirms that these species have been isolated and have followed different evolutionary routes. It should, however, be noted that the results involving the S7 intron 1 marker did not include all the taxa used in the cyt *b* analyses. Nevertheless, the congruence between all molecular markers and the species' diagnostic morphological characteristics (Figure 4.6) provide further evidence that *G. tenuiformis* is distinct from *G. giuris*. The results of the molecular analyses, therefore, clearly contradict the taxonomic decision of a monotypic *G. giuris* species group that includes *G. tenuiformis* (Hoese 1986).

Intraspecific cyt *b* sequence divergences are typically around 1% in fishes (Meyer *et al.* 1990). In a review of the levels of sequence divergence among sister taxa of various fish species, McCune and Loveyjoy (1998) reported a maximum divergence in the cyt *b* region of 5.7% observed in *Mallotus villosus* (Müller, 1776). The maximum divergence observed between sister taxa for gobies was reported as 4.0% between *Gymnogobius* species (Harada *et al.* 2002). Acknowledging the fact that some fish taxa typically exhibit higher levels of sequence divergence than others, the observed levels of sequence divergence between *G. tenuiformis* and the "combined" populations of *G. giuris* (6.5%) is far greater than those recorded for other gobies (see Harada *et al.* 2002). This level of divergence combined with the monophyletic groupings and the morphological differences provide enough evidence for reinstating *G. tenuiformis* as a valid species.

Furthermore, the two morphologically inseparable *G. callidus* lineages coincide well with the hypothesis that *G. callidus* is a species complex (Helen Larson, *pers. comm.*). The lack of morphological differences between the two *G. callidus* clades could be a result of incomplete lineage sorting or the lack of morphological characters observed within the genus *Glossogobius* (Akihito and Meguro 1975). With at least two genetically distinct taxa and a high sequence divergence of 5% (see Chapter 3), its recognition as a single species should be reconsidered. As already reported in Chapter 3, the divergence observed between the

Mozambique lineage and *G. callidus* clade B (1.2%, Figure 4.1) suggested that the former lineage could represent another diverging species of the genus. However, more specimens from Mozambique must be studied and better understanding of interspecific divergence in *Glossogoibius* is required before an informed decision on the status of the Mozambique lineage could be made.

A rather unexpected result of the phylogenetic analyses was the sister group relationships between South African and Australian species in the cyt *b* analyses. These two continents separated at least 37 million years ago (Unmack 2001) and, assuming no gene exchange between the two areas, one would have expected the Australian species to form a strongly supported monophyletic clade separate from the southern African species. Instead, the cyt *b* analyses (Figure 4.2) suggested two different sister group associations between southern African and Australian species. The species of the *G. callidus* group (clades A and B) were the sister group to *Glossogobius* sp., an unidentified Australian species (clade F), while *G. tenuiformis* (clade C) was the sister group of the Australian *G. aureus*. However, sequence divergence between *G. aureus* and *G. tenuiformis* was much lower (3.0%) compared to that of other sister groups (e.g. 6.5% between *G. tenuiformis* (clade C) and African *G. giuris* (clade E) although these species are morphologically distinct and are recognised as separate species (see Akihito and Meguro 1975 for full description of *G. aureus*). It should be noted that even though *G. aureus* was represented by a single individual since the other individual (*G. aureus01*) could be a misidentification), its placement in the tree was highly supported in the Bayesian analysis (100% posterior probability). The southern African (clade D) and the Australian *G. giuris* (clade E) also exhibited a 3% sequence divergence. These two lineages (clade D and E) could thus be representing different species (see Figure 4.1 and 4.2). According to Akihito and Meguro (1975) *G. giuris* is assumed to have a worldwide distribution but may actually consist of more than one species occurring in different areas. While this could not be confirmed by results of the S7 intron 1 gene, the results of the cyt *b* analyses are in agreement with the morphological variability found across the species' distribution (Akihito and Meguro 1975) and imply that *G. giuris* represents a species complex in need of a thorough investigation.

In conclusion, the molecular analyses of this study have shown that previous morphological studies tended to be conservative and underestimated the diversity of *Glossogobius* species. As reported in the literature (see Akihito and Meguro 1975; Hoese and Winterbottom 1979;

Hoese 1986) and from evidence presented in this study it is possible that other known species of this genus may actually represent complexes of species. The information from the molecular data demonstrated that the demotion of *G. tenuiformis* to a junior synonym of *G. giuris* (Hoese and Winterbottom 1979; Hoese 1986) was unjustified and its validity should be re-instated. *Glossogobius callidus* was represented by two genetically divergent lineages, of which one lineage, the KZN lineage (see Chapter 3), should be renamed.

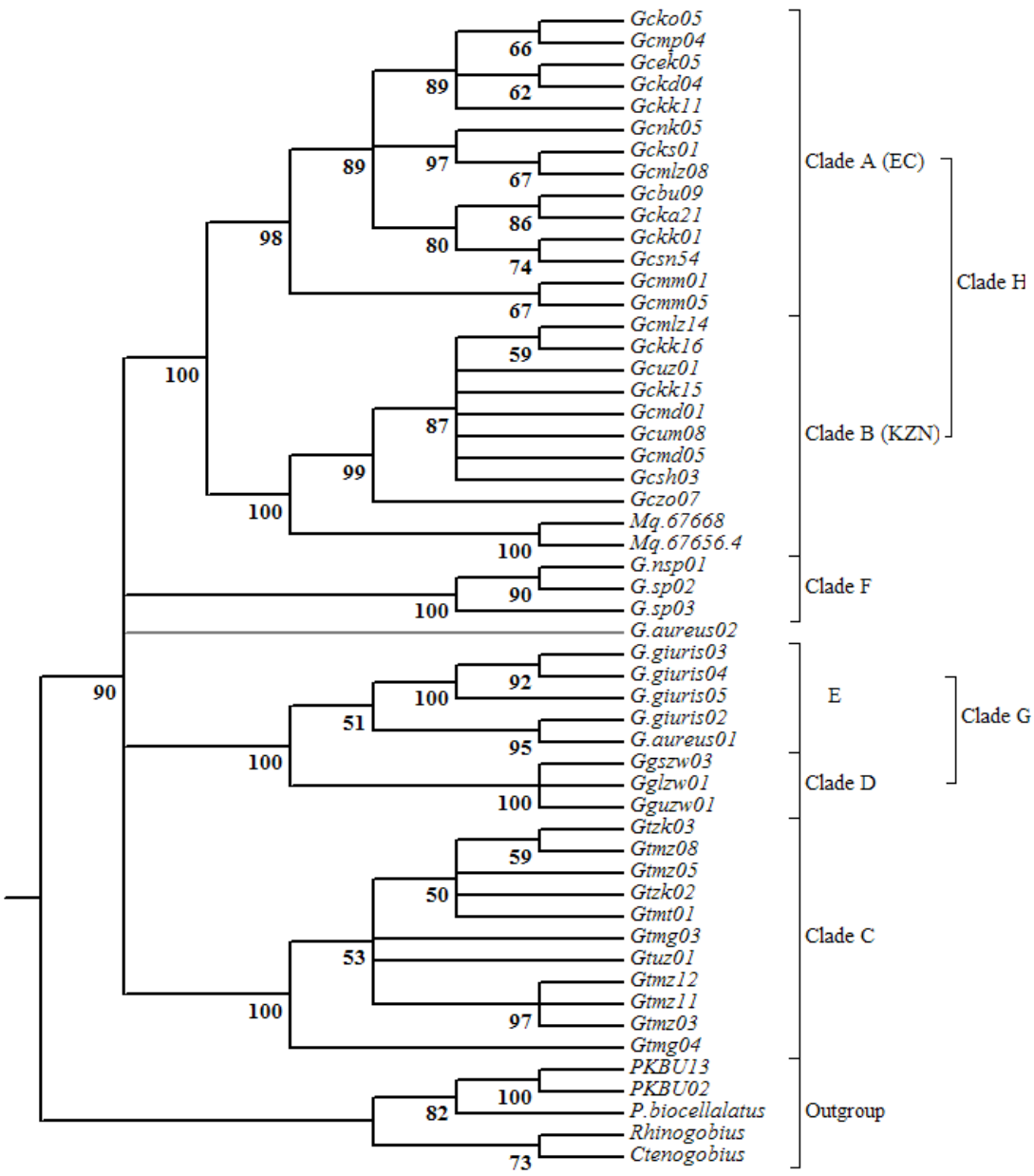


Figure 4.1. A maximum parsimony tree (50% majority rule) of the cytochrome *b* data showing the monophyly and the relationships of the analysed *Glossogobius* species. Clade A = *Glossogobius callidus* sp. 1, clade B = *Glossogobius callidus* sp. 2, clade H = *Glossogobius callidus*, clade C = *Glossogobius tenuiformis*, clade D = African *Glossogobius giuris*, group E = Australian *Glossogobius giuris*, clade G = *Glossogobius giuris* and clade F = Australian *Glossogobius* sp. The first two letters of the taxon names represent species codes while the last two represent the locality names. As outgroups, *Psammogobius* spp. was sequenced during the course of the study while *Rhinogobius* sp. and *Ctenogobius* sp. sequences were downloaded from Genbank.

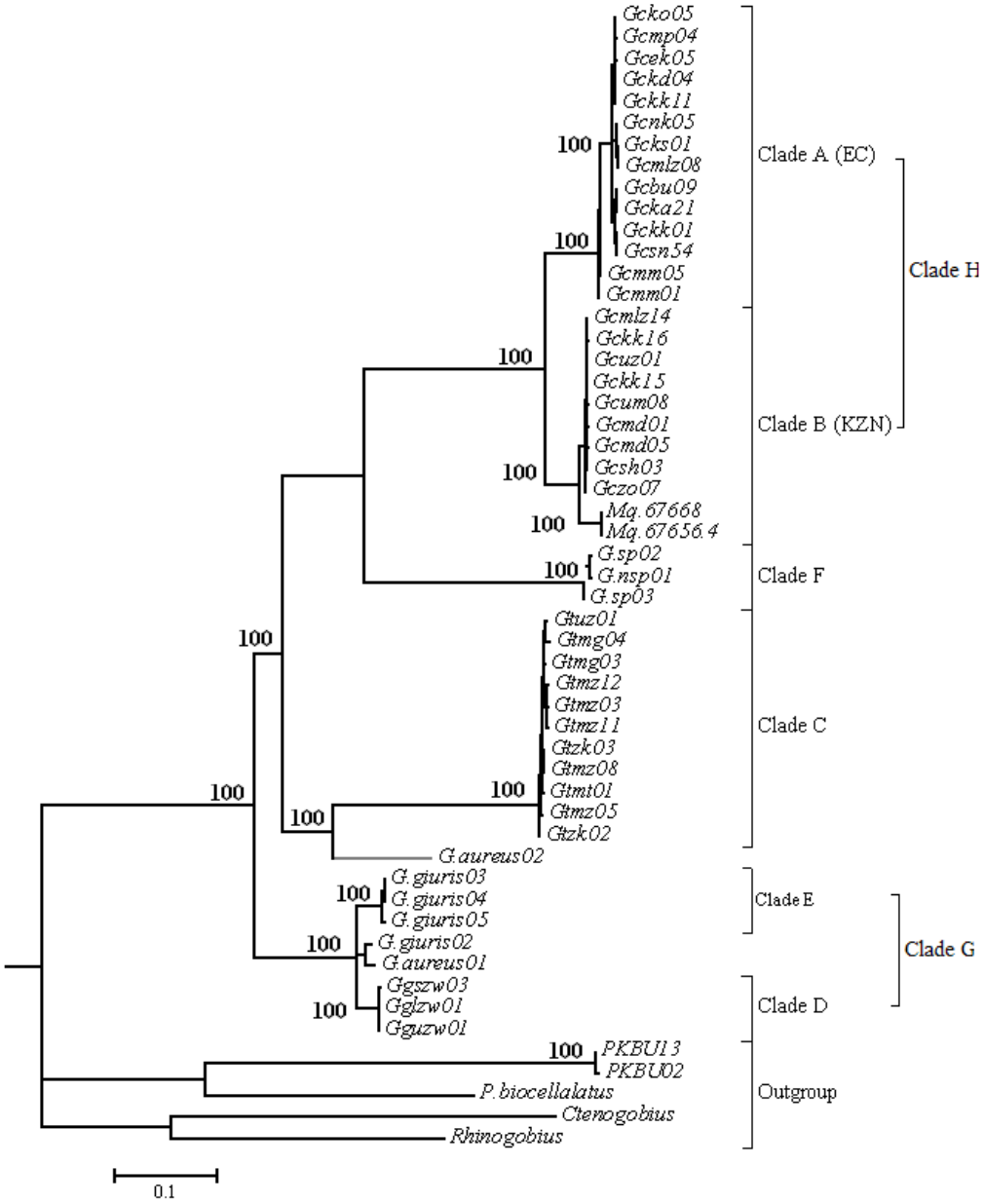


Figure 4.2. A maximum-likelihood phylogram (-lnL = 6744.5) of the cytochrome *b* dataset. Bayesian posterior probabilities are given above each branch. Nodes with less than 95% posterior probabilities are not labelled. Clade A = *Glossogobius callidus* sp. 1, clade B = *Glossogobius callidus* sp. 2, clade H = *Glossogobius callidus*, clade C = *Glossogobius tenuiformis*, clade D = African *Glossogobius giuris*, clade E = Australian *Glossogobius giuris*, clade G = *Glossogobius giuris*, and clade F = Australian *Glossogobius* sp. The first two letters of the taxon names represent species codes while the last two represent the locality names. As outgroups, *Ctenogobius* sp. and *Rhinogobius* sp. sequences were downloaded from Genbank.

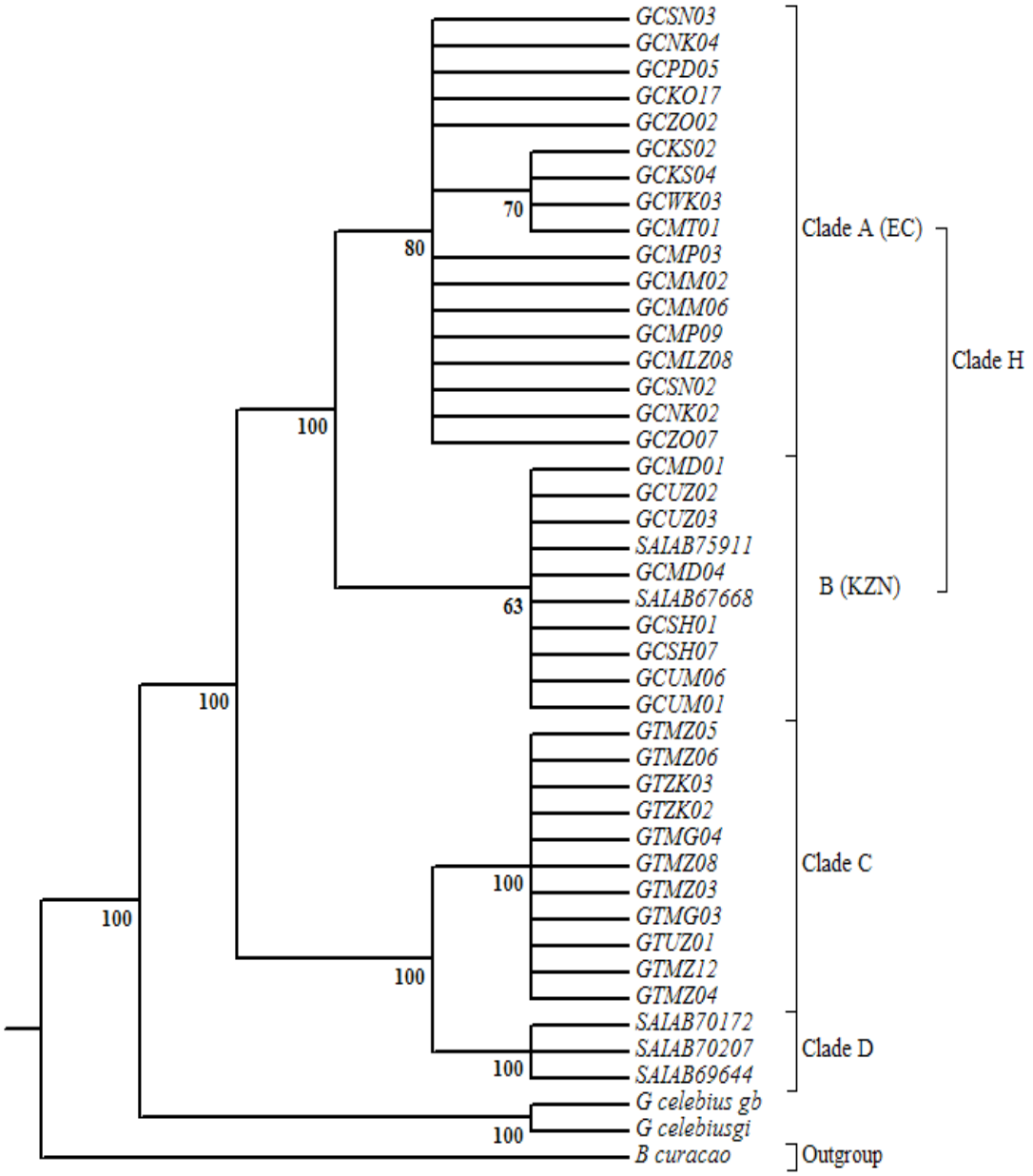


Figure 4.3. A maximum parsimony tree (50% majority rule) of the S7 nuclear intron marker bootstrapped with 1000 replicates. Clade A represents *Glossogobius callidus* sp. 1, group B = *Glossogobius callidus* sp. 2, clade H = *Glossogobius callidus*, clade C = *Glossogobius tenuiformis* and clade D = *Glossogobius giuris.* Specimens taken from the SAIAB DNA collection facility are labelled according to their accessioning codes. The first two letters of the taxon names represent species codes while the last two represent the locality names *Glossogobius celebius gb* and *gi* and the outgroup taxon *B. Curacao* (*Bathygobius curacao*) were downloaded from Genbank.

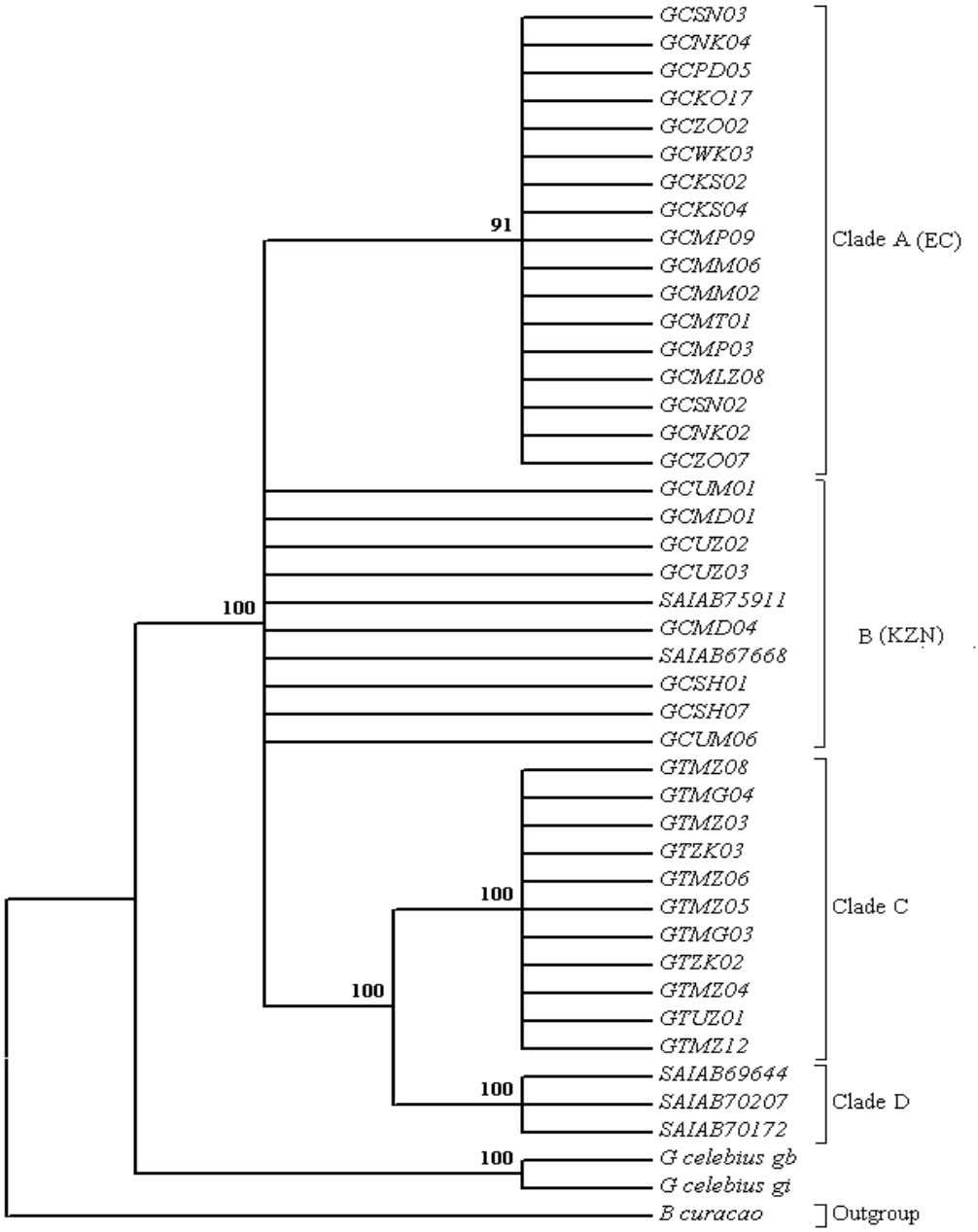


Figure 4.4. Bayesian analysis tree of the S7 intron 1 data without gaps. Clade A represents *Glossogobius callidus* sp. 1, B = *Glossogobius callidus* sp. 2, clade C = *Glossogobius tenuiformis* and clade D = *Glossogobius giuris.* Specimens taken from the SAIAB DNA collection facility are labelled according to their accession codes. The first two letters of the taxon names represent species codes while the last two represent the locality names *Glossogobius celebius gb* and *gi* and the outgroup taxon *B. Curacao* (*Bathygobius curacao*) were downloaded from Genbank.

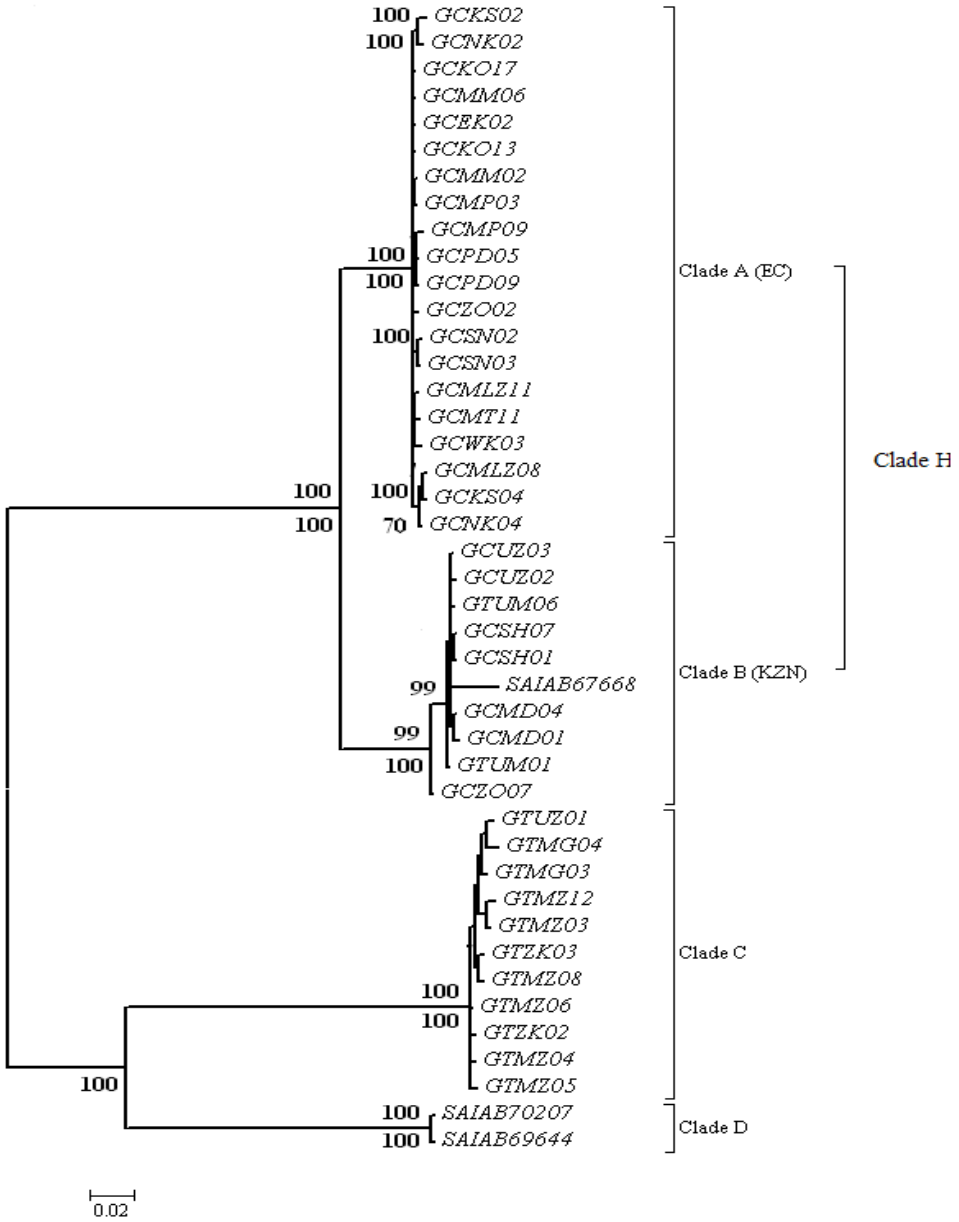
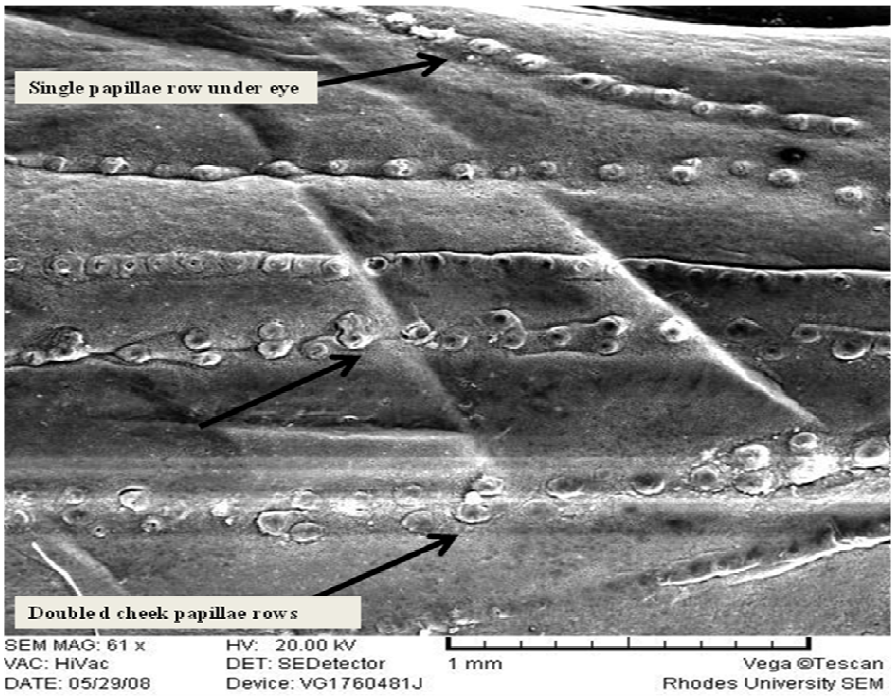


Figure 4.5. A midpoint rooted maximum-likelihood phylogram of the combined dataset. Parsimony bootstrap values are given below branches, and the Bayesian posterior probabilities are given above each branch. Nodes with less than 50% bootstrap support and less than 90% posterior probabilities are not labelled. Clade A represents *Glossogobius callidus* sp. 1, clade B = *Glossogobius callidus* sp. 2, clade H = *Glossogobius callidus,* Clade C = *Glossogobius tenuiformis* and Clade D = *Glossogobius giuris.* Specimens taken from the SAIAB DNA collection facility are labelled according to their accesioning codes. The first two letters of the taxon names represent species codes while the last two represent the locality names.



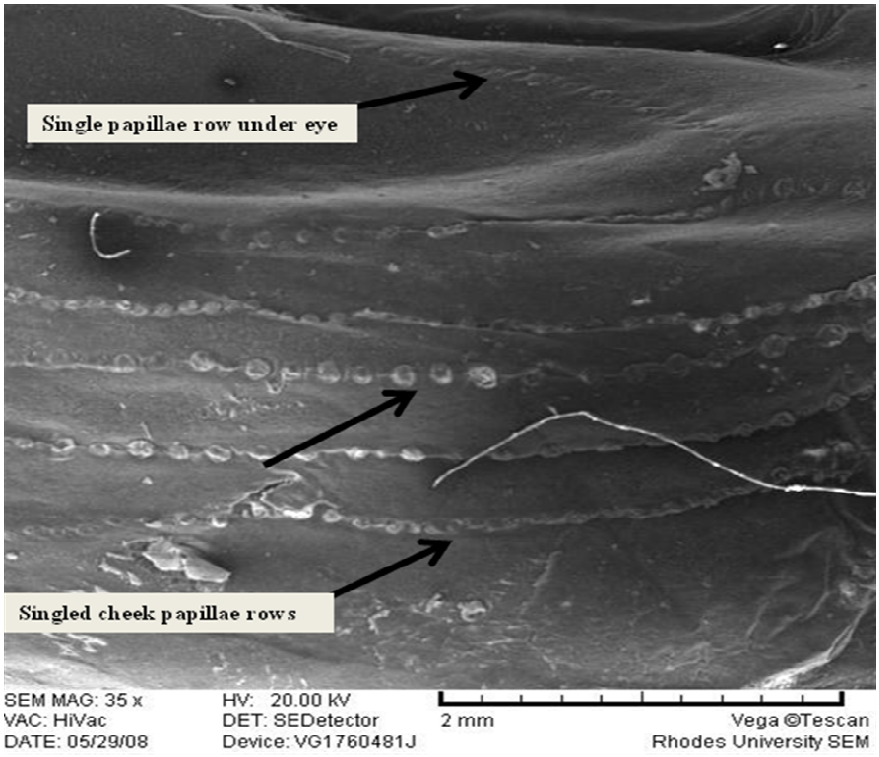


Figure 4.6. Cheek papillae row patterns of *Glossogobius giuris* Hamilton, 1822 (Accession number: SAIAB73983) (top) and *Glossogobius tenuiformis* Fowler, 1934 (specimen number: GTUZ01) (bottom) obtained with the aid of scanning electron microscopy. Note the single papillae row under the eye and the single/double cheek papillae rows.

CHAPTER 5

GENERAL DISCUSSION AND CONCLUSIONS

The primary aim of this study was to provide molecular analyses to assist in the re-evaluation of the taxonomy of the goby genus *Glossogobius* in southern Africa. Information is provided on the population structure and phylogeography of the naturally occurring southern African populations of *G. callidus* using mitochondrial cyt *b* and nuclear S7 intron 1 genes as genetic markers. The same genes were also used to infer phylogenetic relationships among the southern African *Glossogobius* species.

DNA sequence tools now allow us to develop diagnostic characters for identifying species, elucidating their founding populations, and determining whether variants belong to 'cryptic' taxa (Colborn *et al.* 2001; Teske *et al* 2009). The mitochondrial DNA cyt *b* has been applied to many systematic studies of teleost fish where it has proved useful in unravelling phylogenetic relationships (e.g. Harada *et al*. 2002; Haponki and Stepien 2008). This gene was also successfully used in the current study in identifying the phylogeographic structure of the river goby (*G. callidus*) as well as determining the phylogenetic relationships of the southern African flat head *Glossogobius* species (*G. callidus, G. giuris* and *G. tenuiformis*). The mitochondrial cyt *b* provided strong phylogenetic signal, yielding (in most cases) well-resolved and well-supported nodes for intra- and interspecific relationships.

The S7 intron analyses on the other hand had more robustness towards the base and weakness towards the tips of estimated trees, indicating that this gene may be more effective in the estimation of deeper phylogenetic relationships. As reported in the literature the rates of change in nuclear genes may be relatively low compared with other mitochondrial DNA regions (e.g. cyt *b*), such that they have been frequently avoided for interspecific analyses (Simmons and Ocheterena 2000). Nevertheless, there was significant interspecific (Chapter 4) and intraspecific (in terms of haplotype diversity) (Chapter 3) variation observed within this gene, despite the lower number of individuals analysed. This was evident in the levels of haplotype diversity that were relatively high and the significant population structure observed (Φ_{st} = 0.74, F_{ST} = 0.69) within *G. callidus* clades (Chapter 3). Similar intraspecific results have been demonstrated by Chow and Hazama (1998) in detecting significant variation in the first and second introns of S7 introns among yellowfin tuna populations. Their results

suggested that this fragment may be useful in the investigation of genetic structure and phylogenetic relationships among some closely related species. The results of the present study also agree that the first introns of the S7 gene can be effectively used as a molecular marker for studies on interspecific relationships. Although indel results were not significant in this study, as they provided no additional information on the tree topologies and supported the same groups as the de-gapped data set, nuclear indels are important and can provide new information in some other fish taxa (Chow and Hazama 1998; Near *et al.* 2004). A suprising result of the current study was the high number of variable nucleotides (577 (122 + 455)/688 = 86.4%) in this part of the S7 intron 1 sequenced. A previous phylogenetic study on Mormyroidea that used the same gene (Lavoue *et al.* 2003) exhibited only 44% variable nucleotides. Whether this is a characteristic of the species or specific part of the gene should be investigated in future studies.

Using both genes (cyt *b* and S7 intron 1) in their study, Lavoue *et al.* (2003) performed partition homogeneity tests (PHT) or incongruence length difference (ILD) tests to detect significant incongruence (at the $P < 0.05$ level) in their pairwise comparisons of cyt *b*, S7 introns and other used markers. Their S7 intron dataset was incongruent with the mtDNA cyt *b* dataset. However the two datasets were still combined together with other datasets. Even though Bull *et al*. (1993) suggested that incongruent data should not be combined for phylogenetic analyses, the combination of incongruent data can increase the resolution and the support within phylogenetic trees, revealing 'hidden signal' present in the separate datasets (Gatesy *et al*. 1999). One of the objectives of the present study was to assess the suitability of the cyt *b* and S7 intron 1 genes to independently estimate the population structure and phylogeography of *G. callidus*. Therefore, analyses of the two datasets were done separately. Furthermore, since S7 intron 1 could not be successfully sequenced for all the obtained samples of cyt *b*, these datasets were not combined in an attempt to avoid compromising the cyt *b* dataset. Additionally, as indicated by the results of this study, cyt *b* is more effective in estimating intraspecific relationships which was the objective of chapter 3. However, in estimating the phylogenetic relationships of *Glossogobius* species, a marker efficient in the estimation of deeper relationships was required. Due to previous controversial reports (Bull *et al.* 1993; Gatesy *et al.* 1999; Lavoue *et al.* 2003) a partition homogeneity test was not conducted before combining the datasets after separate analyses. Thus in this study the PHT or ILD tests were not used as a criterion to determine the congruency of the separate

datasets, but they were combined to explore the possible "hidden" phylogenetic signal according to Gatesy *et al*. (1999).

It has been suggested that *G. callidus* in South Africa could be a species complex (Helen Larson, *pers. comm*). Both the phylogeographic (see Chapter 3) and phylogenetic analyses (Chapter 4) separated individuals of this species into two well-supported monophyletic clades providing evidence of strong population and phylogeographic structure. Both genetic markers applied in this study indicated a phylogeographic break in the populations associated with the subtropical and warm-temperate biogeographic provinces (Chapter 3). The identified lineages were reciprocally monophyletic and have probably been reproductively isolated. In this respect, it appears that some environmental conditions in each province may have exerted some selection pressure on each phylogroup that resulted in isolation and speciation of the species. Therefore, the results agree with observations that *G. callidus* is a distinct case of a cryptic species (without morphological distinguishing characteristics). There are at least two genetically separate species within the current species and its recognition as a single species should be reconsidered. Similar results were obtained in a study on limpets *Patella granularis* Linnaeus, 1758 (Ridgway *et al*. 1998). Despite morphological similarities between populations of the limpet along the coast of southern Africa, individuals from the northern sites on the east coast represented a gene pool distinct from the west, south and south east coast populations. Ridgway *et al.* (1998) suggested that the lack of gene flow between the populations was due to the physical oceanography and the marked discontinuity of the inshore water characteristics in the Mbashe area which resulted in speciation.

Consequently, *G. callidus* samples forming clade A (see Figure 4.1, EC lineage) would be considered the true *G. callidus* since this clade includes individuals collected from the type locality (Bushman's River) and comprised samples from both freshwater and estuarine localities. Individuals sampled from a restricted area (subtropical region including inland Mozambique) within the range of the current broadly conceived *G. callidus*, all clustered in clade B (KZN lineage), a probable freshwater group. Furthermore, the observed cyt *b* sequence divergence of 5% between these two lineages of *G. callidus* is higher than that expected for intraspecific levels and falls within the range of divergences known for other fish species. It is even higher than values reported between other goby species (e.g. *Gymnogobius* species, see Harada *et al*. 2002).

Glossogobius callidus in the coastal areas is known to breed in the middle and upper reaches of estuaries where it is known to complete its early life history stages (Whitfield 1998; Strydom and Neira 2006). Many localities in the subtropical biogeographic region are generally classified as temporary open systems that are annually disconnected from the sea (see Table 3.1). The build-up of water during the closed phase of such estuaries increases the chance of fish being flushed out to sea when the estuary mouth reopens, particularly when the flooding coincides with the presence of larvae. Flushed individuals may then migrate to other rivers via open estuaries. Although some southern African invertebrate taxa have exhibited intraspecific structuring and gene variations attributed to the effects of isolation due to temporary open systems (see Teske *et al*. 2007a), it seems that *G. callidus* is able to overcome the effect of isolation in the subtropical biogeographic region. Norton (2006) reported that the distance between most estuaries and the coastal rivers in the subtropical province averages 5.3 km while those in the warm temperate average 8.6 km. The closer distances between the estuaries in the subtropical province could be more conducive to successful recruitment into a new locality than in the warm-temperate region, but it is yet to be confirmed. Although the low genetic structuring among individuals of *G. callidus* from the subtropical region could probably be due to these environmental factors, it may also be a result of the small sample size of specimens from this region.

The oceanography of the subtropical region may also play an important role in gene flow among localities. The indication of population expansion may be due, directly or indirectly, to the effect of the ocean's circulation on larval dispersal to and from the estuarine environment after the opening of the temporary open system and flushing of individuals occurs. The flow of the Agulhas Current which follows the continental shelf closely up until the region of Port Alfred (Payne and Crawford 1989; Whitfield 1990) assists in the migration of species along this part of the coast (Whitfield 1998). Shelf waters also mix with the surface waters of the Agulhas Current (Lutjeharms 2005), which assists species to recruit into nearby estuaries. This mixing of water bodies could partially explain the low level of phylogeographic structuring between most of the estuarine subtropical localities within the EC lineage. Similar findings were also reported for other estuarine species, such as *Atherina breviceps* Valenciennes, 1835, in which these factors led to the mixing of populations in the subtropical east coast localities (Norton 2006). The *A. breviceps* results indicated a mixed, population along the South African coastline with high levels of gene flow between estuaries (Norton 2006). Although *G. callidus* is not totally dependent on the estuaries (i.e. can also

live and reproduce in freshwater), its dispersal mechanism in this province may also be similar to that of *A. breviceps* which is dispersed by the actions of flooding and those of the Agulhas Current.

The *G. callidus* analyses also indicated that the riverine populations/localities in the warm-temperate Eastern Cape are genetically grouped with the neighbouring systems, forming geographic groups (groups A, B, C and and clade D) (see Figure 3.3). This observation correlates well with the hypothesis that most freshwater species are generally more genetically structured than estuarine populations. This subdivision can be explained in two ways. First, in addition to the obvious barriers in our river systems, *G. callidus* larvae and individuals in the warm-temperate region mostly occur in the permanently open river estuaries (Strydom and Neira 2006) where the volume of water flowing through the mouth is always at a moderate flow, decreasing chances of individuals being flushed into the sea. This results in the isolation of populations over time. This applies to the studied Eastern Cape localities (see Table 3.1) except for the Kleinemonde localities both of which are temporarily open systems. Second, it can be explained by the sea-level changes and the biogeographic history of the major river systems in the Eastern Cape (such as Bushman's, Great Fish and Sundays) (Swartz *et al.* 2007). A regression of sea level is known to have extended river courses, increasing gradients and providing opportunity for inter-connection of these adjacent systems (Skelton 1980a; Swartz *et al*. 2007). In the long process of evolution the larger rivers have naturally expanded and exchanged drainage with adjacent small rivers (Skelton 1980a). Either way, Skelton (1980b) and Swartz *et al*. (2007) inferred that the recent and historical fragmentation and isolation of river systems was the dominant process that shaped the genetic patterns observed in this biogeographic province. The same pattern of population differentiation and structuring was also observed in the estuarine species *Gilchristella aestuaria* Gilchrist, 1913. The population structure of *G. callidus* along the South African warm-temperate coastline can however to some extent be related to the life history patterns and breeding strategy of *G. aestuaria*. *Gilchristella aestuaria* breeds in the upper reaches of estuaries and completes its entire life cycle in the estuarine environment (Strydom *et al.* 2002). *Gilchristella aestuaria* also avoids being swept out to sea by selecting areas with lower current velocities (Whitfield 1998). This results in relatively few individuals being swept to sea and dispersing through the open ocean (Norton 2006) which leads to the separation of populations and regional lineages. However, flushing events have been reported to result in the movement of *G. aestuaria* from the estuarine to the marine environment,

where the physical oceanography as well as the nature of the South African coastline contributes to the migration of individuals between different regions, especially in the subtropical province (Strydom *et al.* 2002).

To elucidate better structure and to further understand the phylogeography and different evolutionary processes in each biogeographic province, a faster evolving mitochondrial DNA marker (such as the control region and microsatellites), could be applied. Because of their faster evolution and higher mutational rate than cyt *b*, these markers might find structuring not detected using cyt *b* and the S7 intron 1 in the subtropical region. Microsatellites are highly polymorphic and, as such, are often used in population genetic and fingerprinting studies for recently diverged lineages (Avise 2004). Comparative phylogeographic patterns and a better understanding of geo-morphological processes can also help to improve our understanding of the river system evolution in the warm-temperate and the subtropical regions.

From a conservation point of view, the generated data provides evidence of the existence of distinct genetic lineages within *G. callidus.* These lineages may represent historically isolated populations that could be viewed as separate evolutionary units for management purposes. Moritz (1994) proposed Management Units (MUs), which consist of one or more populations showing significant differentiation in their frequencies of nuclear alleles or mitochondrial haplotypes. The Kosi bay/Nkanini lineage (clade D, Figure 3.3) as well as the Mozambique (clade E, Figure 3.3) specimens formed distinct lineages from all other localities (or geographic lineages) for both markers used and had strong nodal support in the estimated phylogeny (see Figure 3.3). Although the sequence divergence of 1.2% between the Mozambique and the rest of the KZN lineage individuals may indicate another unique lineage of the species, more specimens are required to verify this result. Even though the above lineages are morphological inseparable, and lack information concerning their ecological and physiological attributes, these results support the notion that genetically determined phylogroups of widely distributed species, whose distributions closely match the ranges of biogeographic provinces, are likely to be distinct evolutionary significant units (ESU) or MUs (see for examples Rocha *et al.* 2005; Teske *et al.* 2008; Teske *et al.* 2009). In conserving these units, management should aim to preserve adaptive diversity and evolutionary processes across the species geographic range. Therefore, a broader categorization of population distinctiveness should be employed which should not solely be

based on molecular phylogenies and genetic variation of populations but also incorporate ecological data and its concepts (Crandal *et al.* 2000). In order to achieve this, future research should focus on the ecological characteristics of the identified phylogroups from this study.

The findings of this study supports Teske *et al* (2009) 's view on the importance of including sampling sites from the extreme north-east of South Africa (a region that has been poorly studied because of logistical difficulties in reaching sampling sites (Momade and Achimo 2003) and ideally also from Mozambique and beyond, when studying cryptic biodiversity in southern Africa. Although other geographic lineages of *G. callidus* were also observed (e.g. Bushman's, Kariega, see Figure 3.1a), not all of these were genetically distinct or divergent. There was no evidence to suggest that conservation measures, such as genetic supplementation of the species local populations through artificial immigration, are at this time necessary given the current levels of genetic variation within localities and geographic lineages. Moreover, one of the significantly evolving geographic lineages of the species (freshwater Kosi Bay lineage) is within a national heritage site (the iSimangaliso World Heritage Park managed by Ezemvelo KZN Wildlife) and thus may have fewer threats to its survival. Although the Nkanini/Kosi Bay lineage shares haplotypes, Nkanini is apparently not within the iSimangaliso World Heritage Park. Because of this locality's unique genetic makeup, it is worth recommending that this particular locality be included in the already conserved wildlife area to further reduce threats to its survival.

The analyses of generated data resulted in good phylogenetic resolution of southern African flat head goby relationships. All the members of *Glossogobius* used in this study formed a monophyletic group while all the species clades were well supported by high bootstrap and BI values. Although most non-African species of the genus (Table 1.1) were not included in the present analyses, the results suggested that each species of the genus in southern Africa had a sister species occurring in Australia (see Figure 4.2). Even though the Australian species *Glossogobius aureus* (*G. aureus02*) was represented by a single specimen, it emerged as a well supported sister species of the presumed South African endemic *Glossogobius tenuiformis.* Despite the morphological differences observed among representatives of the widely distributed *G. giuris*, the monophyly of representatives of the species from both Australia and Africa was well supported, giving the southern African *G. giuris* a sister group from Australia. However, these two relationships were not supported in the cyt *b* parsimony analysis (Figure 4.1), but were well recorved in the maximum likelihood and strongly

supported by the Bayesian analyses (Figure 4.2). Additionally, *Glossogobius callidus* (sp. 1 and sp. 2) was the sister group of *Glossogobius* sp. (clade F) (Figure 4.2). These sister species relationships were corroborated by the low sequence divergences found between the sisters taxa compared to the divergences between other distant species. The addition of *G. celebius* in the S7 intron 1 data (Figure 4.3) indicated that this Asian and Oceania species could be basal to the monophyletic southern Africa species.

Given the uncertainty associated with current molecular dating procedures, no attempt was made to correlate population divergences to any particular paleogeographic events. Nevertheless, some preliminary biogeographical considerations can be made to explain the observed Australian-African species relationships, bearing in mind the limited scope and taxa included in this study and that the Australian species did not amplify for S7 intron marker. Allopatry (through vicariance) is a plausible mode of speciation for the sister groups of *Glossogobius* species revealed by this study because Australia and Africa were once part of one supercontinent, Gondwanaland. The basal placement of the freshwater Asian species (*G. celebius*) also supports a Gondwana distribution for this genus. The splitting event, about 37 million year ago (Unmack 2001), created the necessary conditions leading to further evolution of the ancestral species along different paths (e.g. random genetic drift) on their respective continents, resulting in cladogenesis.

The observed pattern of interspecific relationships and genetic distances (Figure 5.1) implies that speciation could have taken place in the ancestral southern African *Glossogobius* even before Australia drifted away from southern Africa. It is postulated that this gave rise, either through allopartic or sympatric speciation events, to the ancestral taxa of *G. giuris*, *G. tenuiformis* and *G. callidus* (nodes A and B, Figure 5.1) which in turn split up during the Gondwana breakup (nodes C to E). *Glossogobius callidus* could have diverged into sp. 1 and sp. 2 after Australia moved away (node F). All species then evolved morphologically different although the genetic divergences of the species remained relatively low (e.g. 3% cyt *b* divergence between *G. tenuiformis* and *G. aureus*) in comparison to the "ancestral" species divergence within each continent (e.g. 6.5% divergence between African *G. giuris* and *G. tenuiformis*). The age of the diversification (37 million years ago) for the southern African-Australian *Glossogobius* species broadly coincides with other studies that found evidence for a diversification related to the major climatic variations, geographic isolation and continental drift. For example, some reptiles species of the genus *Coleodactylus* (Sphaerodactylinae,

Gekkota) were estimated to have diverged in the Oligocene, approximately 29.3, ± 4.33 Mya as the effects of continental drift and landscape change (Geurgas *et al.* 2008). Although these reptiles are terrestrial forest inhabitants their speciation event was also a result of rain forest fragmentations (geographic isolation) that could have happened during the drift (Geurgas *et al.* 2008). The hypothesised age of divergence was based on the penalised likelihood and Bayesian dating methods that were considered reliable by these authors. Substantiating these biogeographic interpretations for *Glossogobius* requires a larger study that will incorporate all the known species of this genus. The additional taxa could alter the phylogenetic hypothesis presented in this study (Chapter 4), leading to a different conclusion on the historical biogeography of the genus in southern Africa.

The results obtained during the current study should be viewed against the limitations of the study, namely: (1) most species of the genus could not be included in the dataset (see Table 1.1); and (2) only a subset of taxa used in the cyt *b* analyses were available for the S7 intron 1 analyses. The addition of more taxa and more characters can be expected to substantially improve on the limited picture that emerged from this initial molecular study of the southern African flathead goby evolution and their interspecific relationships. However, the results of this work contribute substantially to refining the phylogeographic structure or cryptic species of *G. calllidus* in the southern African region. It also indicated that the two previously mentioned hypotheses about genetic structuring of a species (i.e. high structure in freshwater populations, and homogeneity in the estuarine populations) do apply to *G. callidus.* The populations of the more estuarine based localities in the subtropical region were homogeneous while those in the localities from the riverine habitats (e.g. Kosi Bay, Nkanini, Great Fish, Sundays, etc.) were phylogeographically structured. The phylogenetic results obtained suggest that the diversity of the genus in southern Africa was underestimated. *Glossogobius tenuiformis* is a well-diverged and diagnosed species from *G. giuris* which is also supported by their morphological characters. Additionally, *G. callidus* clade B (sp. 2) is also indicated as a distinct species from the widely distributed conceived *G. callidus.* These findings show that the number of species in our region has doubled, and that a decision to reinstate *G. tenuiformis* as a valid species is probably justified. New phylogenetic information can now be added to morphological knowledge to produce a more accurate taxonomic framework for *Glossogobius* in our region and elsewhere. Reinterpretation of morphological and developmental data of species of this genus can now be attempted on the basis of these results. In addition, this study is one of several studies (such as Harada *et al.* 2002; Harold *et*

al. 2008; Mavimbela MSc thesis in progress) that provide a better foundation for future species and genus level taxonomic studies of gobies which will, in turn, allow for improved interpretations of the biogeography, ecology, and evolution of this group of fishes.

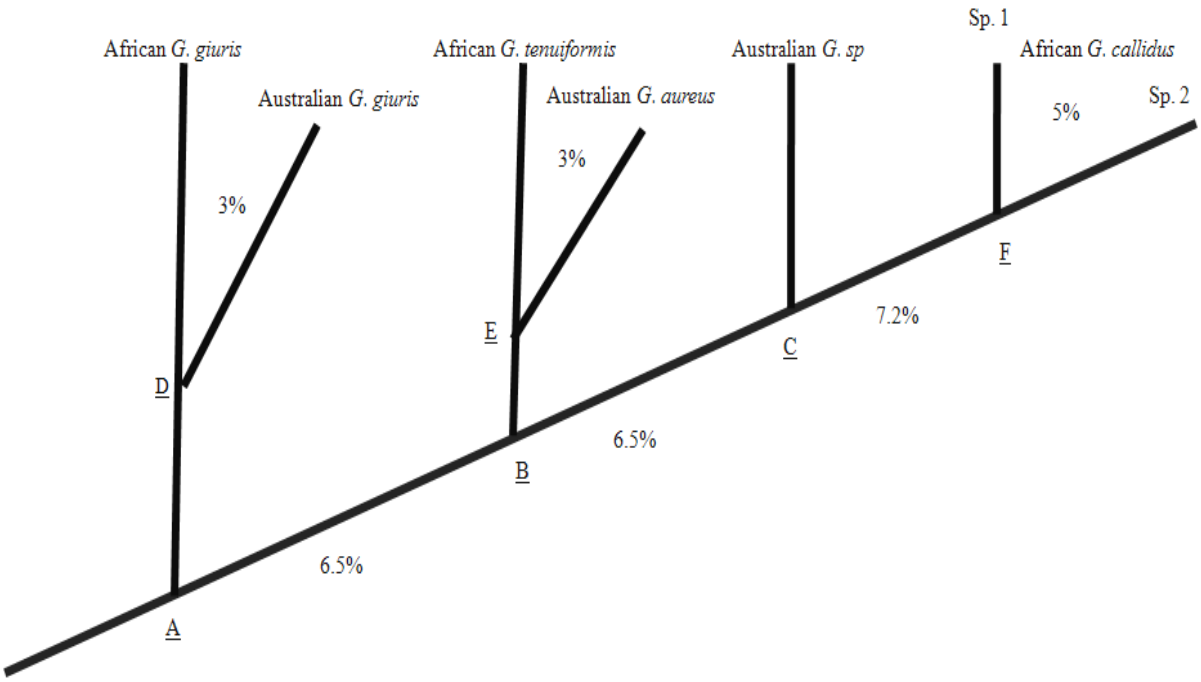


Figure 5.1. The observed pattern of interspecific relationships and genetic distances (mtDNA % divergences) between analysed *Glossogobius* species. The tree is drawn according to the observed results of the cytochrome *b* maximum-likelihood analyses with each single branch representing the species analysed. Each species is labelled with the country of origin. Letters A to F represent the different nodes on the tree, whilst the percentage values are the divergences between the nodes. Australian *G.* sp represents *Glossogobius* sp.; African *G. callidus* is represented by sp. 1 and sp. 2.

REFERENCES

Akihito, P. and Meguro, K. 1975. Description of a new gobiid fish, *Glossogobius aureus*, with notes on related species of the genus. *Japanese Journal of Ichthyology* **22**: 127-142.

Akihito, P., Iwata, A., Kobayashi, T., Ikeo, K., Imanishi, T., Ono, H., Umehara, Y., Hamamatsu, C., Sugiyama, K., Ikeda, Y., Sakamoto, K., Fumihito, A., Ohno, S. and Gojobori, T. 2000. Evolutionary aspects of gobioid fishes based upon a phylogenetic analysis of mitochondrial cytochrome *b* genes. *Gene* **259**: 5-15.

Altukhov, Y.P. 2003. *Intraspecific genetic diversity: monitoring conservation, and management*. PTC Academia, Moscow.

Armstrong, M.H., Braun, E.L. and Kimball, R.T. 2001. Phylogenetic utility of avian ovomucoid introns G: a comparison of nuclear and mitochondrial phylogenies in galliformes. *Auk* **118**: 799-804.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neighel, J.E., Reeb, C.A. and Saunders, N.C. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18**: 489-522.

Avise, J.C. 2000. *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge.

Avise, J.C. 2004. *Molecular markers, natural history, and evolution*. Sinauer Associates Inc, Sunderlands.

Bartlett, S.E. and Davidson, W.S. 1991. Identification of *Thunnus* tuna species by the polymerase chain reaction and direct sequence analysis of their mitochondrial cytochrome *b* genes. *Canadian Journal of Fisheries and Aquatic Science* **48**: 309-317.

Bauchot, M.L., Desoutter, M., Hoese, D.F. and Larson, H.K. 1991. Catalogue critique des types des poissons du Muséum National D'Histoire Naturelle. (Suite) Sous-ordre des Gobioidei. *Bulletin du Muséum National d'Histoire Naturelle, 4th series* **13**: 1-82.

Benzie, J.A.H. 1999a. Major genetic differences between crown-of-thorns starfish (*Acanthaster planci*) populations in the Indian and Pacific Oceans. *Evolution* **53**: 1782-1795.

Benzie, J.A.H. 1999b. Genetic structure of coral reef organisms: Ghosts of dispersal. *American Zoologist* **39**: 131-145.

Birt-Friesten, V.L., Montevecchi, W.A., Gaston, A.J. and Davidson, W.S. 1992. Genetic structure of thick-billed murre (*Uria lomvia*) populations examined using direct sequence analysis of amplified DNA. *Evolution* **46**: 267-272.

Block, B.A. Finnerty, J.R. Stewart, A.F.R. and Kidd, J. 1993. Evolution of endothermy in fish: mapping physiological traits on a molecular phylogeny. *Science* **260**: 10-214.

Blum, M.D. and Tornquist, T.E. 2000. Fluvial responses to climate and sea-level change: a review and look forward. *Sedimentology* **47**: 2-48.

Bowen, B.W., Nelson, W.S. and Avise, J.C. 1993. A molecular phylogeny for marine turtles: trait mapping, rate assessment, and conservation relevance. *Proceedings of the National Academy of Sciences* **90**: 5574-5577.

Brown, W.M., Prager, E.M., Wang, A. and Wilson, A.C. 1982. Mitochondrial DNA sequences of primates: Tempo and mode of evolution. *Journal of Molecular Evolution* **18**: 225-239.

Bull, J.J., Huelsenbeck, J.P., Cunningham, C.W., Swofford, D.L. and Waddell, P.J. 1993. Partitioning and combining data in phylogenetic analysis. *Systematic Biology* **42**: 384–397.

Cantatore, P., Roberti, M., Pesole, G., Ludovico, A., Milella, F., Gadaleta, M.N. and Saccone, C. 1994. Evolutionary analysis of cytochrome *b* sequences in some perciformes: evidence for a slow rate of evolution than in mammals. *Journal of Molecular Evolution* **39**: 589-597.

Carlsson, J., McDowell, J.R., Diaz-Jaimes, P., Carlsson, J.E.L., Boles, S.B., Gold, J.R. and Graves, J.E. 2004. Microsatellite and mitochondrial DNA analyses of Atlantic bluefin tuna

(*Thunnus thynnus thynnus*) population structure in the Mediterranean Sea. *Molecular Ecology* **13**: 3345-3356.

Carvalho, G.R. and Hauser, L. 1998. Advances in the molecular analysis of fish population structure. *The Italian Journal of Zoology* **65**: 21-33.

Chow, S. and Hazama, K. 1998. Universal PCR primers for S7 ribosomal protein gene intron in fish. *Molecular Ecology* **7**: 1247-1263.

Colborn, J., Crabtree, R.E., Shaklee, J.B., Pfeiler, E., Bowen, B.W. 2001. The evolutionary enigma of bonefishes *(Albula spp*): cryptic species and ancient separations in a globally distributed shorefish. *Evolution* **55**: 807-820.

Coronel, J.S, Maes, G.E., Claus, S., Van Damme, P.A. and Volckaert, A.M. 2004. Differential population history in the migratory catfishes *Brachyplatystoma flavicans* and *Pseudoplatystoma fasciatum* (Pimelodidae) from the Bolivia Amazon assessed with nuclear and mitochondrial DNA markers. *Journal of Fish Biology* **65**: 859-868.

Crandall, K.A., Bininda-Emonds, O.R.P., Mace, G.M. and Wayne, R.K. 2000. Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution* **15**: 290-295.

Davies-Coleman, M., Cook, B.A., Harley, E., Leslie, R., Schleyer, M., Cook, P., Brown, C., Cockcroft, V. and Robinson, T. 2000. *Marine Biodiversity Status Report*. Edited by B.D. Durham and J.C. Pauw

Dawson, M.N., Staton, J.L. and Jacobs, D.K. 2001. Phylogeography of the tidewater goby, *Eucyclogobius newberryi* (Teleostei: Gobiidae), in coastal California. *Evolution* **55**: 1167-1179.

Dawson, M.N., Couie, K.D., Barlow, M., Jacobs, D.K. and Swift, C.C. 2002. Comparative phylogeography of sympatric sister species, *Clevelandia ios* and *Eucyclogobius newberryi* (Teleostei: Gobiidae), across the California Transition Zones. *Molecular Ecology* **11**: 1065-1075.

Duran, S., Pascual, M. and Turon, X. 2004. Low levels of genetic variation in mtDNA sequences over the western Mediterranean and Atlantic range of the sponge *Crambe crambe* (Poecilosleri da). *Marine Biology* **144**: 31-35.

Engelbrecht, G.D. and Mulder, P.F.S. 1998. Extremely high genetic differentiation between two populations of the river goby, *Glossogobius callidus* (Smith, 1937). *Water South Africa* **25**: 85-90.

Excoffier, L., Smouse, P.E. and Quattro, J.M. 1992. Analyses of molecular variance inferred from metric distance among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479-491.

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

Field, K.G., Olsen, G.J., Lane, D.J., Giovannoni, S.J., Ghilselin, M.T., Raff, E.C., Pace, N.R. and Raff, R.A. 1988. Molecular phylogeny of the animal kingdom. *Science* **239**: 748-753.

Flemming, B.W. and Hay, E.R. 1988. *Sediment distribution and dynamics on the Natal continental shelf. In: Schumann, E.H. (ed). Coastal ocean studies off Natal, South Africa.* Springer, New York.

Fowler, H.W. 1934. Fishes obtained by Mr. H. W. Bell-Marley chiefly in Natal and Zululand in 1929 to 1932. *Proceedings of the Academy of Natural Sciences of Philadelphia* **86**: 405-514.

Francisco, S.M., Viera, M.N. and Almada, V.C. 2006. Genetic structure and historical demography of the shanny *Lipophrys pholis* in the Portuguese coast based on mitochondrial DNA analysis. *Molecular Phylogenetics and Evolution* **39**: 288-292.

Fu, Y.X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.

Fu, Y.X. and Li, W.H. 1999. Coalescing into the 21st century: an overview and prospects of coalescent theory. *Theoretical Population Biology* **56**: 1-10.

Fujita, M., Engstron, T.N., Starkey, D.E. and Shaffer, N.B. 2004. Turtle phylogeny: insights from a novel nuclear intron. *Molecular Phylogenetics and Evolution* **31**: 1031-1040.

Funk, D.J. and Omland, K.E., 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology and Systematics* **34**: 397–423.

Gadagkar, S., Resenberg, M.S. and Kurmar, S. 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *Journal of Experimental Zoology* **304**: 64-74.

Gaffney, P.M. 2000. Molecular tools for understanding population structure in Antarctic species. *Antarctic Science* **12**: 288-296.

Gatesy, J., O'Grady, P. and Baker, R.H. 1999. Corroboration among data sets in simultaneous analysis: hidden support for phylogenetic relationships among higher level artiodactyls taxa. *Cladistics* **15**: 271–313.

Geurgas, S. R., Rodrigues, M.T. and Moritz, C. 2008. The genus *Coleodactylus* (Sphaerodactylinae, Gekkota) revised: A molecular phylogenetic perspective. *Molecular Phylogenetics and Evolution* **49:** 92-101.

Giribet, G. and Wheeler, W. 1999. On Gaps. *Molecular Phylogenetics and Evolution* **13**: 132-143.

Goren, M. 1996. *A review of the Southern African Gobiid fish genus Caffrogobius Smitt 1900.* JLB Smith Institute of Ichthyology. Grahamstown, South Africa.

Grant, W.S. and Bowen, B.W. 1998. Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *Journal of Heredity* **89**: 415-426.

Greenwood, P.H. 1994. Gobiid fishes of the genus *Glossogobius* in the Limpopo system: a first record and a range extension of *Glossogobius callidus* (Smith, 1937). *Southern African Journal of Aquatic Science* **20**: 88-92.

Haponki, A.E. and Stepien, C.A. 2008. Molecular, morphological, and biogeographic resolution of cryptic taxa in the Greenside darter *Etheostoma blennioides* complex. *Molecular Phylogenetics and Evolution* **49**: 69–83.

Harada, S., Jeon, S.R., Kinoshita, I., Tanaka, M. and Nishida, M. 2002. Phylogenetic relationship of four species of floating gobies (*Gymnogobius*) as inferred from partial mitochondrial cytochrome b gene sequences. *Ichthyological Research* **49**: 324-332.

Harold, A.S., Winterbottom, R., Munday, P.L. and Chapman, R.W. 2008. Phylogenetic relationships of Indo-Pacific coral gobies of the genus *Gobiodon* (Teleostei: Gobiidae), based on morphological and molecular data. *Bulletin of Marine Science* **82**: 119-136.

Harpending, H.C. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* **66**: 591-600.

Harris, S.A. 1996. *Larval fish assemblages of selected estuarine and coastal systems in KwaZulu-Natal, South Africa*. PhD thesis, University of Zululand.

Harris, S.A. and Cyrus, D.P. 2000. Comparison of larval fish assemblages in three large estuarine systems, KwaZulu-Natal, South Africa. *Marine Biology* **137**: 527-541.

Harrison, I.J. 1989. Specialisation of the gobioid palatopterygoquadrate complex and its relevance to gobioid systematics. *Journal of Natural History* **23**: 325-353.

Harrison, T.D. 2004. Physico-chemical characteristics of South African estuaries in relation to the zoogeography of the region. *Estuarine, Coastal and Shelf Science* **61**: 73-87.

Hedges, S.B., Moberg, K.D. and Maxson, L.R. 1990. Tetrapod phylogeny inferred from 18S and 28S ribosomal RNA sequences and a review of the evidence for amniote relationships. *Molecular Biology and Evolution* **7**: 607-633.

Heemstra, P.C. and Heemstra, E. 2004. *Coastal fishes of southern Africa*. NISC / SAIAB, South Africa.

Hewitt, G.M. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907- 913.

Hewitt, G.M. 2004. Genetic consequences of climatic oscillations in the Quatenary. *Philosophical Transactions of the Royal Society of London* **359**: 183-195.

Hey, J. and Harris, E. 1999. Population bottlenecks and patterns of human polymorphism. *Molecular Biology and Evolution* **16**: 1423-1426.

Hillis, D.M. and Moritz, C. 1990. *Molecular Systematics*. Sinauer Press, Sunderland.

Hoese, D.F. and Winterbottom, R. 1979. A new species of Lioteres (Pisces: Gobiidae) from KwaZulu-Natal, with a revised checklist of South African Gobies and comments on the generic relationships and endemism of Western Indian Ocean Gobioids. Royal Ontario Museum. *Life Science Occassional Papers* **31**: 1-13.

Hoese, D.F. 1986. Gobiidae. In: *Smith's Sea Fishes*. Smith M.M. and P.C. Heemstra (eds). Macmillan, South Africa.

Hoese, D.F. and Gill, A.C. 1993. Phylogenetic relationships of eleotridid fishes (Perciformes: Gobioidei). *Bulletin of Marine Science* **52**: 415-418.

Hou, Z., Fu, J. and Li, S. 2007. A molecular phylogeny of the genus *Gammarus* (Crustacea: Amphipoda) based on mitochondrial and nuclear gene sequences. *Molecular Phylogenetics and Evolution* **45**: 596-611.

Huelsenbeck, J.P. and Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**: 754-755.

Inoue, J.G., Miya, M., Tsukamoto, K. and Nishida, M. 2000. Complete mitochondrial DNA sequence of the Japanese sardine, *Sardinops melanostics*. *Fisheries Science* **66**: 924-932.

Irwin, D.M., Kocher, T.D. and Wilson, A.C. 1991. Evolution of the cytochrome *b* gene in mammals. *Journal of Molecular Evolution* **32**: 128-144.

Islam, N.M. 2004. Eco-biology of the freshwater goby, *Glossogobius giuris* (Hamilton) of the River Padma in relation to its fishery: A review. *Journal of Biological Sciences* **4:** 780-793

Karl, S.A., Bowen, B.W. and Avise, J.C. 1992. Global population genetic structure and male-mediated gene flow in the green turtle (*Chelonia mydas*): RFLP analyisis of anonymous nuclear loci. *Genetics* **131**: 163-173.

Kocher, T.D. and Stepien, C.A. 1997. *Molecular Systematics of fishes*. Academic press Limited, London.

Kuriiwa, K., Hanzawa, N.M Yoshino, T., Kimura, S. and Nishida, M. 2007. Phylogenetic relationships and hybridisation in rabbit fishes (Teleostei: Siganidae) inferred from mitochondrial and nuclear DNA analyses. *Molecular Phylogenetics and Evolution* **45**: 69–80.

Ladoukakis, E.D., Saavedra, C., Magoulas, A. and Zouros, E. 2002. Mitochondrial DNA variation in a species with two mitochondrial genomes: the case of *Mytilus galloprovincialis* from the Atlantic, the Mediterranean and the Black Sea. *Molecular Ecology* **11**: 755-769.

Lavoue, S. Sullivan, J.P. and Hopkins, C.D. 2003. Phylogenetic utility of the first two introns of the S7 ribosomal protein gene in African electric fishes (Mormyroidea: Teleostei) and congruence with other molecular markers. *Biological Journal of the Linnean Society* **78**: 273-292.

Lutjeharms, J.R.E. 2005. The coastal oceans of south-eastern Africa. In*: The sea*. Robinson, A.R. and K. Brink, (eds). Chicago University Press, Chicago.

Martin, A.P. and Palumbi, S.R. 1993. Protein evolution in different cellular environments: cytochrome b in sharks and mammals. *Molecular Biology and Evolution* **4**: 873-891.

Mavimbela, C. (in progress). *Molecular systematic of the South African gobiid fishes of the genus Caffrogobius* (*Smith, 1900*), *and the intraspecific genetic variation of Caffrogobius gilchristi* (*Boulenger, 1900*). MSc dissertation, University of Limpopo.

McCune, A.R. and Loveyjoy, N.R. 1998. The relative rate of sympatric and allopatric speciation in fishes: Endless Forms. In*: Species and Speciation.* Howard, D.J. and S.H. Berlocher, (eds). Oxford University Press, New York.

Meyer, A., Kocher, T.D., Basasibuaki, P. and Wilson, A.C. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* **247**: 550-553.

Meyer, A. 1993. Evolution of mitochondrial DNA in fishes. In*: Biochemistry and molecular biology of fishes. Vol. 2.* Hochachka, P.W and T.P. Mommsen, (eds). Elsevier Science Publishers, New York.

Meyer, A. 1994. Shortcomings of the cytochrome *b* gene as a molecular marker. *Trends in Ecology and Evolution* **9**: 278- 290.

Mindel, D.P., Sorenson, M.D, Dimcheff, D.E, Hasegawa, M., Ast, J.C. and Yur, T. 1999. Interoidal relationships of birds and other reptiles based on whole mitochondrial genome. *Systematic Biology* **48**: 138-152.

Momade, F. and Achimo, M. 2003. Dune geomorphology in Maputaland, Mozambique. *Geological Society of America* **35**: 47-57.

Moritz, C., Dowling, T.E. and Brown, W.M. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics* **18**: 269-292.

Moritz, C. 1994. Defining evolutionary significant units for conservation. *Trends in Ecology and Evolution* **9**: 401-375

Moritz, C. 1995. Uses of molecular phylogenies for conservation. *Philosophical Transactions: Biological Science* **349**: 113-118.

Morrison, C.L., Lemarié, D.P., Wood, R.M., King, T.L. 2006. Phylogeographic analyses suggest multiple lineages of *Crystallaria asprella* (Percidae: Etheostomatinae). *Conservation Genetics* **7**: 129–147.

Müller, K. 2005. SeqState - primer design and sequence statistics for phylogenetic DNA data sets. *Applied Bioinformatics* **4**: 65-69.

Mullis, K.B. 1990. The unusual origin of the polymerase chain reaction. *Scientific American* **262:** 56-63.

Near, T.J., Bolnick, D.I. and Wainwright, P.C. 2004. Investigating phylogenetic relationships of sunfishes and black basses (Actinopterygii: Centrarchidae) using DNA sequences from mitochondrial and nuclear genes. *Molecular Phylogenetics and Evolution* **32**: 344-357.

Neethling, M., Mathee, C.A., Bowie, R.C.K. and Von der Heyden, S. 2008. Evidence for panmixia despite barriers to gene flow in the southern African endemic, *Caffrogobius caffer* (Teleostei: Gobiidae). *BioMed Central Evolutionary Biology* **8:** 325

Nei, M. and Li, W. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceeding of the National Academy of Science of the USA* **76**: 5269-5273.

Nei, M. and Tajima, F. 1980. DNA polymorphism detectable by restriction endonucleases. *Genetics* **97**: 145-163.

Neigel, E.J. 1997. A comparison of alternative strategies for estimating gene flow from genetic markers. *Annual Review of Ecology and Systematics* **28**: 105-128.

Neira, F.J., Potter, I.C. and Bradley, J.S. 1992. Seasonal and spatial changes in the larval fish fauna within a large temperate Australian estuary. *Marine Biology* **122**: 1-16.

Neira, F.J. and Potter, I.C. 1994. The larval fish assemblage of the Normalup-Walpole Estuary, a permanently open estuary on the southern coast of Western Australia. *Australian Journal of Marine and Freshwater Research* **45**: 1193-1207.

Nelson, J.S. 2006. *Fishes of the world*, 4th edition. John Wiley and Sons Inc, New Jersey.

Norton, O.B. 2006. *Population structure of two estuarine fish species, Atherina breviceps (Pisces: Atherinidae) and Gilchristella aestuaria (Pisces: Clupeidae), along the southern African coastline.* MSc dissertation. Rhodes University.

Oosthuizen, A., Jiwaji, M. and Shaw, P. 2004. Genetic analysis of the *Octopus vulgaris* populations on the coast of South Africa. *South African Journal of Science* **100**: 603-607.

Palumbi, S.R. and Baker, C.S. 1994. Contrasting population structure from nuclear intron sequences and mitochondrial DNA of humpback whales. *Molecular Biology and Evolution* **11**: 435-436.

Palumbi, S.R. 1997. Molecular biogeography of the pacific. *Coral Reefs* **16**: 47-52.

Payne, A.I.L. and Crawford, R.J.M. 1989. *Oceans of life off southern Africa*. Vlaeberg Publishers, South Africa.

Penzo, E., Gandolfi, G. Bargelloni, L. and Colombo, L. 1998. Messianian salinity crisis and the origin of freshwater lifestyle in Western Mediterranean gobies. *Molecular Biology and Evolution* **15**: 1472- 480.

Polzin, T. and Daneschmand, S.V. 2003. On Steiner trees and minimum spanning trees in hypergraphs. *Operations Research Letters* **31:** 12-20.

Porat, N. and Botha, G. 2008. The luminescence chronology of dune development on the Maputaland coastal plain, southeast Africa. *Quaternary Science Reviews* **27**: 1024-1046.

Posada, D. and Crandal, K.C. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics Applications Note* **14**: 817-818

Posada, D. and Crandal, K.C. 2001. Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution* **16**: 37-45.

Pusey, B., Kernnard, M. and Arthington, A. 2004. *Freshwater fishes of north-Eastern Australia.* CSIRO publishing. Australia.

Qi, D., Guo, S., Zhao, X., Yang, J. and Tang, W. 2007. Genetic diversity and historical population structure of *Schizopygopsis pylzovi* (Teleostei: Cyprinidae) in the Qinghai-Tibetan Plateu. *Freshwater Biology* **52**: 1090-1104.

Quattro, J.M., Greig, T.W., Coykendall, D.K., Bowen, B.W. and Baldwin, J.D. 2002. Genetic issues in aquatic species management: the shortnose sturgeon (*Acipenser brevirostrum*) in the southeastern United States. *Conservation Genetics* **3**: 155–166.

Riddle, B.R., Hefner, D.J. and Alexander, L.F. 2000. Phylogeography and systematics of the *Peromyscus eremiscus* species group and the historical biogeography of North American warm regional deserts. *Molecular Phylogenetics and Evolution* **17**: 146- 160.

Ridgway, T.M., Sterwart, B.A., Branch, G.M. and Hodgson, A.N. 1998. Morphological and genetic differentiation of *Patella granularis* (Grastropoda: Patellidae): recognition of two sibling species along the coast of southern Africa. *The Zoological Society of London* **245**: 317-333.

Rocha, L.A., Robertson, D.S., Roman, J. and Bowen, B.W. 2005. Ecological speciation in tropical reef fishes. *Proceedings of the Royal Society of London B: Biological Science* **272**:573–579.

Roe, K.J., Harris, P.M. and Mayden, R.L. 2002. Phylogenetic relationships of the genera of North American sunfishes and basses (Percoidei: Centrarchidae) as evidenced by the mitochondrial cytochrome *b* gene. *Copeia* **4**: 897-905.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X. and Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496-2497.

Schlotterer, C. 2001. Genealogical inference of closely related species based on microsatellites. *Genetics research* **78**: 209-212.

Schluter, D. 2000. *The ecology of adaptive radiation*. Oxford University Press, New York.

Schneider, S. and Excoffier, L. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* **152**: 1079-1089.

Schneider, S., Roessli, D. and Excoffier, L. 2000. *Arlequin ver 2.000: A software for population genetics data analysis.* Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Schumann, E.H. 1988. Physical oceanography off Natal. In: *Coastal Ocean studies off Natal, South Africa*. Schumann, E.H. (eds). Lecture Notes on Coastal and Estuarine Studies vol 26, Spriger Berlin.

Simmons, M.P. and Ocheterena, H. 2000. Gaps as characters in sequence-based phylogenetic analysis. *Systematic Biology* **49**: 369-381.

Simmons, M.P., Ocheterena, H. and Carr, T.G. 2001. Incorporation, relative homoplasy, and effects of gap character in sequence-based phylogenetic analyses. *Systematic Biology* **50**: 454-462

Simonsen, K.L., Churchill, G.A. and Aquadro, C.F. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**: 413-429.

Sink, K.J., Branch, G.M., Harris, J.M. 2005. Biogeographic patterns in rocky intertidal communities in KwaZulu-Natal, South Africa. *African Journal of Marine Science* **27**: 81-96.

Skelton, P.H. 1980a. *Systematics and biogeography of the redfin Barbus species (Pisces: Cyprinidae) from southern Africa*. PhD thesis. Rhodes University.

Skelton, P.H. 1980b. Aspects of freshwater fish biogeography in the Eastern Cape. *The Eastern Cape Naturalist* **24**: 17-22.

Skelton, P.H. 1993. *A complete guide to the freshwater fishes of Southern Africa*. 1st edition. Southern Book Publishers (Pty) Ltd, South Africa.

Skelton, P.H. 2001. *A complete guide to the Freshwater Fishes of Southern Africa*. 2nd edition Struik Publishers, South Africa.

Slade, R.W., Moritz, C., Hoelzel. A.R. and Burton H.R. 1998. Molecular population genetics of the southern elephant seal *Mirounga leonine. Genetics* **149**: 1945-1957.

Slatkin, M. and Maddison, W.P. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* **123**: 603-613.

Smith, J.L.B. 1960. Fishes of the family Gobiidae in South Africa. *Ichthyological Bulletin Rhodes University* **18**: 299-314.

Stefanni, S. and Thorley, J.L. 2003. Mitochondrial DNA phylogeography reveals the existence of an Evolutionary Significant Unit of the sand goby *Pomatocshistus minutus* in the Adriatic (Eastern Mediterranean). *Molecular Phylogenetics and Evolution* **28**: 601-609.

Steiner, C., Tilak, M., Douzery, E. and Catzeflis, F. 2005. New DNA data from a transthyretin nuclear intron suggest an Oligocene diversification of living South American opossums (Marsupialia: Didelphidae). *Molecular Phylogenetics and Evolution* **35**: 363-379.

Stepien, C.A. and Faber, J.E. 1998. Population genetic structure, phylogeography and spawning philopatry in walleye (*Stizostedion vitreum*) from mitochondrial DNA control region sequences. *Molecular Ecology* **7**: 1757-1769.

Strydom, N.A., Whitfield, A.K. and Paterson, A.W. 2002. Influence of altered freshwater flow regimes on abundance of larval and juvenile *Gilchristella aestuaria* (Pisces: Clupeidae) in the upper reaches of two South African estuaries. *Marine and Freshwater Research* **53**: 431 - 438.

Strydom, N.A. Whitfield, A.K. and Wooldridge, T.H. 2003. The role of estuarine type in characterizing early stage fish assemblages in warm temperate estuaries, South Africa. *African Zoology* **38**: 219-243.

Strydom, N.A. and Neira, F. 2006. Description and ecology of larvae of *Glossogobius callidus* and *Redigobius dewaali* (Gobiidae) from temperate South African estuaries. *African Zoology* **41**: 240-251.

Stuckenberg, B.R. 1969. Effective temperature as an ecological factor in southern Africa. *Zoological Africa* **4**: 145-197

Sunnucks, P. 2000. Efficient markers for population biology. *Trends in Ecology and Evolution* **15**: 199-203.

Swartz, E.R., Skelton, P.H. and Bloomer, P. 2007. Sea level changes, river capture and the evolution of population of the Eastern Cape and fiery redfins (*Pseudobarbus afer* and *Pseudobarbus phlegethon*) across multiple river systems in South Africa. *Journal of Biogeography* **34**: 2086-2099.

Swofford, D.L. 1999. *PAUP*. Phylogenetic Analysis Using Parsimony (*and other Methods).* Sinaeur Associates, Sunderland, MA.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585-595.

Templeton, A.R., Routman, E. and Phillips, C.A. 1995. Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the Tiger Salamander, *Ambystoma tigrinum. Genetics* **140**: 767-782.

Teske, P.R., Cherry, M.I. and Matthee, C.A. 2003. Population genetics of the endangered Knysna seahorse, *Hippocampus capensis*. *Molecular Ecology* **12**: 1703-1715.

Teske, P.R., McQuaid, C.D., Froneman, C.D. and Barker, N.P. 2006. Impact of marine biogeographic boundaries on phylogeographic patterns of three South African estuarine crustaceans. *Marine Ecology Progress Series* **314**: 283-293.

Teske, P.R., Froneman, P.W., Baker, N.P. and McQuaid, C.D. 2007a. Phylogeographic structure of the caridean shrimp *Palaemon peringueyi* in South Africa: further evidence for intraspecific genetic units associated with marine biogeographic provinces. *African Journal of Marine Science* **29**: 253-258.

Teske, P.R., Papadopulos, I., Zardi, G.I., McQuaid, C.D., Griffiths, C.L., Edkins, M.T. and Barker, N.P. 2007b. Implications of life history for genetic structure and migration rates of five southern Africa coastal invertebrates: planktonic, abbreviated and direct development. *Marine Biology* **153**: 697-711.

Teske, P.R., Papadopoulos, I., Newman, B.K., Dworschak, P.C., McQuaid, C.D. and Barker, N.P. 2008. Oceanic dispersal barriers, adaptation and larval retention: an interdisciplinary assessment of potential factors maintaining a phylogeographic break between sister lineages of an African prawn. *BioMed Central Evolutionary Biology* **8**:341.

Teske, P.R., Winker, H., McQuaid, C.D. and Barker, N.P. 2009. A tropical/subtropical biogeographic disjunction in southeastern Africa separates two Evolutionarily Significant Units of an estuarine prawn. *Marine Biology* **156**: 1265-1275.

Thacker, C.E. and Cole, K.S. 2002. Phylogeny and evolution of the gobiid genus *Coryphopterus*. *Bulletin of Marine Science* **70**: 837-850.

Thacker, C.E. 2003. Molecular phylogeny of the gobioid fishes (Teleostei: Perciformes: Gobioidei). *Molecular Phylogenetics and Evolution* **26**: 354-368.

Thacker, C.E. and Hardman, M.A. 2005. Molecular phylogeny of basal gobioid fishes: Rhyacichthyidae, Odontobutidae, Xenisthmidae, Eleotridae (Teleostei: Perciformes: Gobioidei). *Molecular Phylogenetics and Evolution* **37**: 858-871.

Thompson, J.D., Gibson, T.J., Plemniak, F., Jeanmougin, F. and Higgins, D.G. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **22**: 4673-4680.

Turbeville, J.M., Pfeifer, D.M., Field, K.G. and Raff, R.A. 1991. The phylogenetic status of Arthropods, as inferred from 18S rRNA sequences. *Molecular Biology and Evolution* **8**: 669-689.

Tweddle, D. 2007. First record of the river goby *Glossogobius callidus* (Smith, 1937) from Malawi. *African Journal of Aquatic Science* **32**(1)**:** 89.

Tzeng, C.S., Lin, Y.S., Lin, S.M., Wang, T.Y. and Wang, F.Y. 2006. The phylogeography and population demography of selected freshwater fishes in Taiwan. *Zoological Studies* **45**: 285-297.

Unmack, P.J. 2001. Biogeography of Australian freshwater fishes. *Journal of Biogeography* **28**: 1053-1089.

Unseld, M., Beyerrmann, B., Brandt, P. and Hiesel, R. 1995. Identification of the species origin of highly processed meat products by mitochondrial DNA sequences. *PCR Methods Application* **4**: 241-243.

Van der Elst, R.P. 1988. *A guide to the common sea fishes of southern Africa*. Struik, South Africa.

Von der Heyden, S., Lipinski, M.R. and Mathee, C.A. 2007. Mitochondrial DNA analysis of the Cape hakes reveal an expanding, panmictic population for *Merluccius capansis* and population structuring for mature fish in *Merluccius paradoxus*. *Molecular Phylogenetics and Evolution* **42**: 517-527.

Vrijenhoek, R.C. 1998. Conservation genetics of freshwater fish. *Journal of Fish Biology* **53**: 394-412.

Wade, M.J., McKnight, M.L. and Shaffer, H.B. 1994. The effects of kin-structured colonisation on nuclear and cytoplasmic genetic diversity. *Evolution* **48**: 1114-1120.

Wang, H.Y., Tsai, M.P., Dean, J. and Lee, S.C. 2001. Molecular phylogeny of gobioid fishes (Perciformes: Gobioidei) based on mitochondrial 12S rRNA sequences. *Molecular Phylogenetics and Evolution* **20**: 390-408.

Whitfield, A.K. 1990. Life-history styles of fishes in South African estuaries. *Environmental Biology of Fishes* **28**: 295-308.

Whitfield, A.K. 1998. Biology and ecology of fishes in southern African estuaries. *Ichthyological Monograph of the J.L.B Smith Institute of Ichthyology* No. **2**.

Whitfield, A.K. 2000. *Available scientific information on individual southern African estuarine systems.* WRC Report No. 577/3/00. Water Research Commission, SouthAfrica. Pp. 217.

Young, N.D. and Healy, J. 2003. GapCoder automates the use of indels characters in phylogenetic analysis. *BioMed Central Bioinformatics* **4**: 1-6.

Zhang, D.X. and Hewitt, G.M. 2003. Nuclear DNA analysis in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* **12**: 563-584.