Genetic diversity in the anabantids *Sandelia capensis* and *S. bainsii*: A phylogeographic and phylogenetic investigation

by

Heidi Roos

Submitted in fulfillment of part of the

requirements for the degree

Master of Science

in the

Department of Genetics

Faculty of Natural & Agricultural Sciences

University of Pretoria

Pretoria

Supervisor: Prof. P. Bloomer

Co-supervisors: Dr. J.A. Cambray

Mr. N.D. Impson

Prof. C.Z. Roux

January 2004

DECLARATION

I declare that the work presented in this thesis is my own, unaided work, except as acknowledged in the text. It is submitted as the requirement for the degree Master of Science at the University of Pretoria, Pretoria, South Africa. It has not been submitted before for any degree or examination at this or any other university.

…….

Acknowledgements

I am very grateful to my supervisor, Dr. Paulette Bloomer for her guidance and support and also for providing good working facilities for the duration of my project. I thank Dr. Jim Cambray and Mr. Dean Impson for providing essential samples and for reviewing my thesis at such short notice. I thank the National Research Foundation (NRF) and Prof. C. Z. Roux for providing bursaries. I am grateful to WWF South Africa (Table Mountain Fund) and Western Cape Nature Conservation Board (WCNCB) who provided funding for my project and to the NRF who awarded grants to Prof. C. Z. Roux and Prof. P. Bloomer. The Mazda Wildlife Fund provided a vehicle for part of my field trips. Thank you to Ernst Swartz, Michael Cunningham, Dr. Jim Cambray and Dean Impson with whom I had many fruitful discussions. I thank Ernst Swartz and Michael Cunningham for many useful comments on earlier drafts. Thanks to Lucy Scott for creating the maps. I would not have been able to collect the samples for my project without Ernst Swartz who showed me the ropes in the field. Isa-Rita Russo and Roelien van Niekerk also helped me with the collection of samples. Thank you. I would like to thank all the farmers that allowed me to sample on their lands. I am grateful to Roger Bills who provided me with hard-to-get samples, some of them from foreign countries. Thank you for your support, Vincent, for reviewing my thesis and for letting me complete a large part of this work on your computer. To Loray, for your most precious guidance and support, I cannot thank you enough. To my best friend Lisel, you have always been there for me, through the good and the bad times. I am extremely grateful to my parents, who have made all of this possible for me, and for their unconditional love and support.

Abstract

Genetic diversity in the anabantids *Sandelia capensis* and *S. bainsii*: A phylogeographic and phylogenetic investigation

by

Heidi Roos

Supervisor: Prof. Paulette Bloomer

Molecular Ecology and Evolution Program

Department of Genetics

University of Pretoria

Pretoria

Co-supervisors:

Dr. Jim A. Cambray

Albany Museum

Rhodes University

Grahamstown

Mr. N. Dean Impson

Western Cape Nature Conservation Board

Stellenbosch

Prof. Carel Z. Roux

Department of Genetics

University of Pretoria

Pretoria

This study concerns the phylogeography of *Sandelia capensis* and *S. bainsii* (family Anabantidae), two freshwater fish species endemic to Cape coastal rivers of South Africa.

The mitochondrial DNA (mtDNA) cytochrome b gene (*S. capensis*) and control region (*S. capensis and S. bainsii*) were used as genetic markers. *Sandelia capensis* has a wide distribution, and occurs in most river systems of the Cape Floristic Region (CFR). Therefore, by studying the genetic variation within the species we could investigate the drainage history of this region. Two major historically isolated lineages were identified within this species, one comprising west coast populations and the other south coast populations. Genetically unique lineages were also identified within each of these two major clades. The split between the two major clades dated back to the Pliocene, while divergence times for lineages within them dated back to the Pleistocene. River capture events and sea level changes in the CFR played a major role in shaping the genetic variation that we observe within *S. capensis* today. *Sandelia bainsii* is restricted to a few Eastern Cape coastal rivers, and is classified as endangered. Within this species two divergent clades were identified, a Great Fish/Kowie lineage and a Buffalo/Gulu lineage. *Sandelia capensis* is also of conservation importance since certain populations are declining as a result of many different threats. In identifying these genetically unique lineages, certain areas (rivers/populations) could be prioritized for conservation management. The two major lineages identified within each of the two species should be conserved as separate units and not be intermixed. Two genetically very unique populations identified within *S. capensis*, the Heuningnes and Diep, are under severe pressure and should also be prioritized for management. In addition, a preliminary phylogenetic study was performed on the Anabantidae using mtDNA 16S rRNA sequences. This was done to investigate the phylogenetic relationships between the two *Sandelia* species (classified as sister taxa) and also their relationships with the other anabantids. The phylogenetic relationships between the anabantids were largely unresolved, probably due to an ancient radiation.

Contents

List of Tables

Chapter 2

Chapter 3

List of Figures

Chapter 2

Chapter 1: General Introduction

The Cape Floristic Region (CFR) or Cape Floral Kingdom of South Africa is the smallest and most distinctive of the six floral kingdoms in the world and is situated at the southern tip of Africa. It stretches in an L-shape down the west coast of South Africa to Cape Town and eastwards towards Port-Elizabeth (van Wyk & Smith 2001). The CFR is a priority area for conservation nationally and internationally since it is a center of biodiversity and endemism and many of the animal and plant species of this region are threatened with extinction. Large parts of the natural habitat have already been destroyed and degraded, mainly because of agriculture, infestation by alien invasive plants and urbanization. Sixteen of its 19 indigenous primary freshwater fish species are endemic or near endemic, and nine of these are endangered or critically endangered (Impson *et al.* 2002; IUCN 2003).

Sandelia capensis Cuvier, 1831, and *S. bainsii* Castelnau, 1861, are two freshwater fishes endemic to the Cape region (the CFR and Eastern Cape respectively) of South Africa. The two species belong to the family Anabantidae (Osteichthyes; Perciformes), and although they are very distinct from each other, they have been classified as sister taxa (Norris 1994). The Anabantidae (climbing perches) forms part of the suborder Anabantoidei. Members of the family Anabantidae are known as the labyrinth fishes, since they have an accessory air-breathing organ, the labyrinth, which is located in a chamber above the gills (Skelton 1993). The labyrinth enables these fishes to survive in tropical swamps and streams (habitat of all genera of the family, except for *Sandelia*) where oxygen levels are often low. In *Sandelia* the air-breathing organ is reduced, possibly due to Cape coastal drainages not being poorly oxygenated (Barnard 1943). There are four genera within the Anabantidae: the Asian *Anabas* and three African genera, *Ctenopoma, Microctenopoma* and *Sandelia* (Cambray 1997a). The anabantids have a disjunct distribution, with *Anabas* occurring in south-east Asia, *Ctenopoma* and *Microctenopoma* in west central, central and south-eastern Africa and *Sandelia* in Cape

coastal rivers of South Africa (Norris 1994). There is a total of 12 species in the genus *Ctenopoma*, of which *Ctenopoma multispine* Peters, 1844 (manyspined climbing perch) is the only one found in southern Africa. It occurs in the rivers of coastal Mozambique and northern KwaZulu-Natal, the Okavango, upper and lower Zambezi, and Kafue drainages (Skelton 1993). It is also found in the Quanza system in Angola and the southern tributaries of the Congo (Skelton 1993). There are 14 species in the genus *Microctenopoma* (Elsen 1976). *Microctenopoma intermedium* Pellegrin, 1920 (blackspot climbing perch) is found in the Okavango, upper and lower Zambezi and Kafue Rivers. It also occurs in the St. Lucia basin in KwaZulu-Natal and the southern tributaries of the Zaire system (Skelton 1993). *Sandelia* forms an unusual part of the ichthyofauna of the Cape, since it is isolated in the southern tip of Africa and separated from the rest of its family (Skelton 1993). The two *Sandelia* species have a disjunct distribution, with *S. bainsii* restricted to eastern Cape coastal systems, while *S. capensis* is restricted to the CFR and is found in coastal drainages from the Coega River near Port Elizabeth in the east to the Langvlei River in the west (Cambray 1997a).

Species biology

In general, *S. capensis* (or the Cape Kurper) is golden-brown in colouration, with dark markings on some of the scales (Harrison 1952). There is a tendency for darker shading on the back and lighter shading ventrally (Harrison 1952). The species is relatively easily recognized by dark streaks radiating from the eyes backwards to the edge of the preoperculum (Harrison 1952). Although not well studied, *S. capensis* shows extensive morphological variation across its distribution range (Barnard 1943): not only are there large differences in the shape of the body, but also considerable variation in markings and colour (Harrison 1952). For instance, *S. capensis* from muddy and opaque rivers were more silvery in colour with a greenish tinge, and in some instances, these specimens also had a pinkish tinge around the axil of the pectoral fin (Harrison 1952). Specimens that are almost black in colour have also been documented (Harrison 1952). Boulenger suggested that specimens from Port Elizabeth, because of an increase in the number of dark streaks on the head and dark spotting, are a distinct species (*S. vicinus*) (Harrison 1952). Barnard (1943) however believed this to be only a colour variety of *S. capensis*. There are four

variations on the typical body shape of *S. capensis* (see Harrison 1952). In a large amount of material examined, specimens collected from the south-western Cape were documented to have dorsal spines varying in number from 12-14 and the anal spines from six to eight (Barnard 1943). In specimens collected further East (George-Port Elizabeth) the dorsal spines varied from 13-14, while the anal spines were eight and in certain specimens, seven. *Sandelia capensis* and *S. bainsii* are distinguished from each other by the following: *S. bainsii* has small scales along the bases of its dorsal and anal fins, while *S. capensis* lacks such small scales; *S. capensis* has pronounced black markings on its head and body which is not found in *S. bainsii* (Skelton 1993).

The Cape Kurper can reach a length of up to 200 mm SL, is very hardy and can survive in a wide range of habitats (Harrison 1952). It prefers slow-flowing water with plant cover (Skelton 1993). Its accessory breathing organ, although reduced, allows it to utilize oxygen in the atmosphere to a certain extent. It can therefore survive in water with low oxygen content and for small periods of time outside of water. It is an omnivorous feeder feeding on insects, other invertebrates and smaller fish (Skelton 1993). It is a very aggressive fish and hides underneath rocks or in plants, waiting for its prey. Breeding time is during summer and males guard the eggs in a breeding territory (Skelton 1993). Breeding males turn darker and have intensely black fins and distinctive black striping, especially on the head (Cambray 1990; Cambray 2004, in press). Males do not develop contact organs (Cambray 1980).

Sandelia bainsii or the Eastern Cape Rocky is olive or grey brown with scattered black flecks and spots and a single bar behind the eye (Cambray 1996). The species is the largest of the anabantids, reaching a length of up to 260 mm SL (Cambray 1997b). As the name implies, it lives in rocky streams where it hides beneath the rocks and waits for its prey (insects, crabs and smaller fish) to pass by (Cambray 1996). Breeding takes place in summer (October to February) and males guard the eggs and prepare a spawning site (Cambray 1996). Breeding males turn darker and develop contact organs (Cambray 1996).

Phylogeography

The phylogeography of each of the two *Sandelia* species was studied using mitochondrial DNA (mtDNA) markers. Phylogeography concerns the study and explanation of DNA molecular variation of a species in space and time; it is the genetic composition and structure of extant populations (Avise 2000). Many examples of phylogeographic studies on freshwater fish using mtDNA data exist in the literature (e.g. Bermingham & Avise 1986; Nedbal & Philipp 1994; Machordom & Doadrio 2001; Perdices *et al*. 2002; Perdices *et al*. 2003). Mitochondrial DNA has a higher rate of evolution than single copy nuclear DNA (about 1-10× faster) and is therefore appropriate for use in population-level or intraspecific questions (Meyer 1994). This fast rate of evolution of mtDNA has made it the marker most often used in population and evolutionary biology (Meyer 1994). Other characteristics of mtDNA make it a good molecular system for phylogenetic analysis (Avise 1987). Mitochondrial DNA is easily isolated and displays a straightforward mode of inheritance (maternally inherited in most species) since it does not undergo homologous recombination (Avise 1987). One limitation of mtDNA as a genetic marker is heteroplasmy (when two or more genotypes coexist within the same individual), however, this has a small impact, since such cases are unusual (Avise 1987). The molecule's maternal mode of inheritance only allows the inference of maternal gene-flow, therefore a nuclear marker should also be used to determine paternal gene-flow patterns. The genetic markers used in this study were the mtDNA control region and cytochrome b gene. The control region appears to control mtDNA replication and RNA transcription and is not transcribed (Avise 1987). The protein product encoded by the cytochrome b gene plays a role in energy production that is essential for survival (Avise 1987). Since the control region does not code for RNA or protein, it evolves at a faster rate than cytochrome b, in which mutations might influence protein function and will therefore be visible for natural selection (Meyer 1994). The control region may therefore be more useful for population-level questions, such as determining gene-flow among populations.

Conservation aspects

Although *S. capensis* is not classified as endangered, many populations are declining (Baille & Groombridge 1996). *Sandelia bainsii* is classified as endangered (IUCN 2003) and several populations are on the verge of extinction (Cambray 1997a). Many different threats are responsible for the imperilled conservation status of the freshwater fishes of the Cape; these are predation and competition (for food and space) from alien invasive fishes such as bass, trout, tilapia and catfish, habitat destruction through poor agricultural practices, excessive water extraction, poor water quality, pesticides and alien invasive plants (see Skelton 1993 and Impson *et al.* 2002). Therefore, this study was also of conservation importance, since the identification of genetically unique populations of threatened endemic fishes would contribute to the prioritization of certain rivers for conservation management.

It was not the aim of this study to take part in the debate surrounding conservation units and their identification (e.g. Moritz 1994a; Pennock & Dimmick 1997; Waples 1998; Bowen 1999; Crandall *et al.* 2000). I decided to follow the sometimes opposing views of two investigators (Moritz 1994b; Crandall *et al.* 2000) as a guideline in suggesting appropriate management actions for *S. capensis* and *S. bainsii*. I used Moritz' (1994b) definition for a conservation unit: the evolutionarily significant unit or ESU. Moritz (1994b) defined ESUs (a term first coined by Ryder 1986) as populations or sets of populations that have been historically isolated and therefore have distinct evolutionary potential. According to Moritz (1994b), ESU's should be reciprocally monophyletic for mtDNA alleles and show significant divergence for nuclear allele frequencies. The ESU is used in a variety of contexts by investigators throughout the world and is recognized under the Endangered Species Act in the USA (Moritz 1994b). In the classification of the ESU (Moritz 1994b), no information concerning adaptive differences or ecological data is incorporated. Crandall *et al.* (2000) state the importance of categorizing distinct populations on the basis of ecological and genetic exchangeability. Individuals that are ecologically exchangeable have similar life history traits, demographic characteristics, morphologies and ecological requirements (Crandall *et al.* 2000). Individuals are

genetically exchangeable when "there is ample gene flow between populations" (Crandall *et al.* 2000).

Aims of this project

Chapter 2 describes a phylogeographic study on *S. capensis*. The wide distribution and morphological variation observed within *S. capensis* makes it a good study species from a phylogeographic perspective. Knowledge of the genetic variation within *S. capensis* would not only provide more insight into the history of CFR drainages (the historical connections between them), but also allow the identification of units for conservation management. This knowledge will be important for the identification of specific rivers in the CFR that should be prioritized for conservation management, since it is impossible to conserve all rivers and populations within the CFR. Typically, such river systems will contain genetically unique populations that are under threat from habitat destruction, alien invasive fishes, water extraction, and other threats in the CFR. Genetic structuring within the species would be expected, since Cape coastal drainages are currently isolated. In this study I wanted to identify significant phylogeographic breaks and the processes (river capture, sea level changes, migration within systems) that played an important role in shaping the genetic patterns within *S. capensis*.

In Chapter 3 the genetic diversity within the endangered Eastern Cape Rocky was studied in order to identify if there were any unique conservation units that need to be managed separately. With this we wanted to identify the processes that have occured within the species and that should be allowed to continue in future. The species faces extinction if drastic management plans are not implemented. One of the management plans considered for the species is a captive breeding program. Chapter 3 also describes a preliminary phylogenetic study on the anabantids to test whether *S. capensis* and *S. bainsii* are indeed sister species and to investigate their phylogenetic relationships with the other anabantids.

Chapter 2: Phylogeography of the Cape Kurper *Sandelia capensis* using mitochondrial DNA data

2.1 Abstract

A widely distributed freshwater fish, *Sandelia capensis,* was used to investigate the drainage history of the Cape Floristic Region (CFR) of South Africa and to identify distinct regions that should be prioritised for conservation management. The levels of mitochondrial DNA (mtDNA) control region and cytochrome b sequence variation found in *S. capensis* were among the highest reported for freshwater fish. Two major historically isolated lineages were identified, one comprising south coast populations and the other west coast populations. In addition, divergent lineages were identified within each major phylogenetic group: Heuningnes, Breede/Goukou; Gourits (Bos and Langtou Rivers); Gamtoos/Maitlands and Baakens/Maitlands lineages within the southern group and Verlorenvlei/Langvlei; Berg and Diep lineages within the west coastal group. A molecular clock dated the split between the two major lineages to be of early-late Pliocene age (4–2 million years ago). Divergence times estimated for lineages within each of the two major groups spanned the entire Pleistocene epoch. Therefore, changing climatic and environmental conditions in the last four million years (Pliocene–Pleistocene times) seem to have played a major role in shaping the genetic variation observed in *S. capensis* today. The drainage history of the CFR had played a major role in shaping these complex genetic patterns. Sea level regressions and various river capture events at different times in the past had made it possible for fish to colonize new systems and expand their distribution ranges. Following these dispersal events was fragmentation and isolation when rivers became separated after river capture and rise in sea level. The major south and west coast lineages are distinct conservation units and should be protected as such: no translocations should be allowed between them. The distinct Heuningnes and Diep populations should be prioritized for conservation management since they are under severe threat, mainly from alien fishes and habitat destruction.

2.2 Introduction

Sandelia capensis is an anabantid and is closely related (a sister species) to *S. bainsii* (Norris 1994). It is smaller in size than *S. bainsii*: *S. capensis* reaches a length of up to 200 mm SL while *S. bainsii* can reach a length of up to 260 mm SL (see Chapter 1). The Cape Kurper is widespread across the CFR, but certain populations are threatened. There is extensive morphological variation across the species' distribution range; not only is there variation in markings and colour, but also in the shape of the body (see Chapter 1).

The coastal drainages of the CFR of South Africa (to which *S. capensis* is endemic) are separated by land and sea, which form barriers to the dispersal of primary freshwater fishes such as *S. capensis*. The freshwater fishes of the CFR are only able to disperse to new drainages when contact between systems has been established. Two geological phenomena would have played an important role in the historical connections among drainage basins of the CFR. The first is river capture, which takes place when a river breaks through a mountain range and captures the headwaters of another drainage. The second would have been river confluence after a drop in sea level (see Skelton 1980a). River capture is more the result of chance events and not directly determined by, for instance, climatic changes. Changes in sea level, on the other hand, are intricately linked to climatic changes. Skelton (1980a) suggested that river capture processes were mainly responsible for the dispersal of a common ancestor of the *Pseudobarbus* species (a freshwater endemic to the Cape and a co-distributed species of *S. capensis*) throughout the rivers of the Cape Fold Mountains. Examples of such historical river capture events are between the Breede and Gourits systems, between the Gourits drainage and the Groot River of the Gamtoos system and between the Keurbooms and Gourits drainages (Figure 1).

The dispersal of fish into many of the smaller coastal systems is thought to have taken place during periods of low sea level, since most of these rivers do not penetrate the watershed of the coastal mountain chains (Skelton 1980a). Sea level is influenced by two events: tectono-eustacy and glacio-eustacy (Seisser & Dingle 1981). Tectono-eustacy refers to a change in volume of the ocean because of the uplift or subsidence of the Mid-

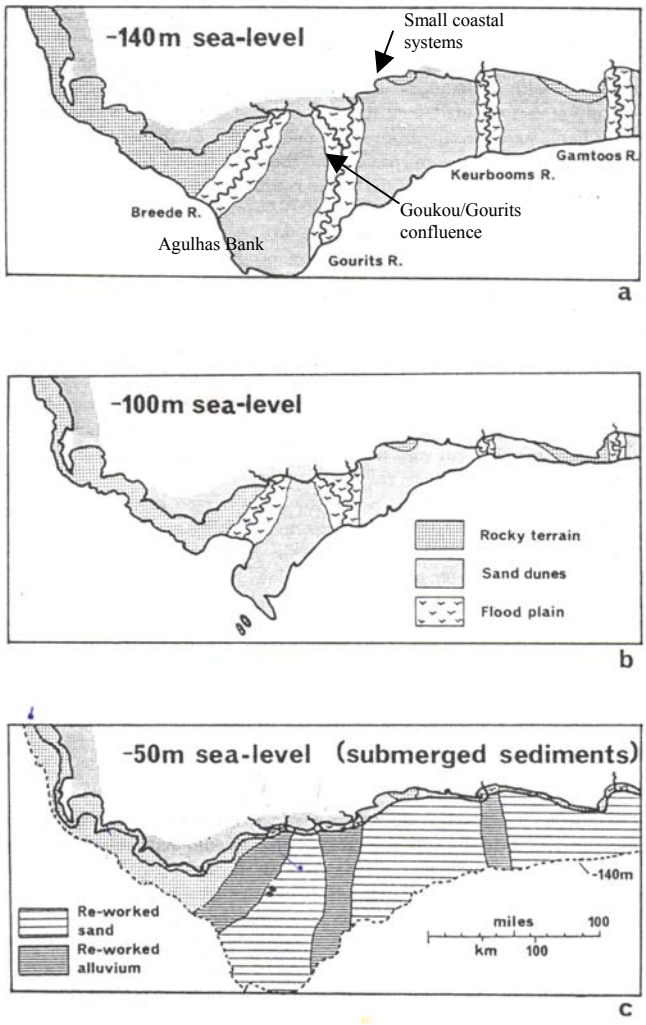

Figure1 Paleogeographies of the Agulhas Bank for selected sea-levels (a. –140; b. –100; c. –50) during the Pleistocene, showing superficial sediment types (Dingle &Rogers 1972).

Oceanic Ridge system. This causes a volume decrease or increase respectively of the ocean-basin capacity, which results in water spilling over the continents (transgression) or decreasing water level (regression) (Seisser & Dingle 1981). Changes in sea level since the Miocene were primarily caused by changes in the volume of the polar ice-caps (glacio-eustacy), resulting in the addition (interglacials) or withdrawal (glacials) of large volumes of water from the sea (Hendey 1983). A major sea level regression has been postulated for the Miocene-Pliocene boundary, considered to date back 3.5-7 million years (m.y.) before present (BP) by some and 4-5 m.y. BP by others (Seisser & Dingle 1981). The last major sea level regression cannot be dated accurately, but investigators consider it to be late Pliocene-early Pleistocene in age. During this time, sea level fell up to 80m on the east coast and 210m on the west coast of South Africa (Seisser & Dingle 1981). These glacio-eustatic sea level changes occurred with greater frequency in the Pleistocene, with interglacials as short as 10 000 years. During the Quaternary interglacials the sea level was similar to what it is today, or in some instances a little higher (Hendey 1983).

There were at least four major glacial periods throughout the Pleistocene epoch during which sea level stands were much lower than at present (Dingle & Rogers 1972). All of these glacial periods appear similar in magnitude (Tankard 1975). These drops in sea level resulted in the exposure of large areas of the continental shelf, especially on the Agulhas Bank, a triangular-shaped continental margin at the southern tip of Africa (Figure 1). During the last glacial maximum (approximately 18 000 years ago), the sea level fell so severely (about 120m below its present level) that the coastal plain widened from 50 – 200km to the south of where the Breede River's mouth is today (Hendey 1983). Comparable exposures of the continental shelf probably took place during glacial maxima earlier in the Pleistocene (Hendey 1983). These drops in sea level would have caused coastal rivers to extend their courses over the exposed continental shelf. The low gradient, especially on the south coast (see Chapter 1) would have caused rivers to meander and to be very wide. Certain drainages (for example the Goukou and Gourits drainages; the small south coastal systems between Mossel Bay and Knysna) would have

had a shared confluence during these times (Figure 1), making it possible for *S. capensis* to disperse between systems.

Skelton (1980b) groups the south coastal rivers into three categories: the great river systems that extend well inland and originate on the Great Escarpment (for example the Gamtoos and Gourits), drainages that originate on "intermediate" mountain ranges such as the Cape Fold belt (for example the Kowie and Keurbooms), and the small coastal drainages such as the Maitlands and Duiwe. Drainage systems that fall within the first two categories are thought to be ancient, their origin believed to date back to the initial isolation of the subcontinent (early post-Gondwana). Rivers in the last category are much younger and believed to have originated in the Tertiary, Pleistocene or even as recent as the Holocene (Skelton 1980a). Fish from the oldest drainages (categories one and two) would have dispersed to and colonized the younger systems, either during times of low sea level and/or during river capture event(s).

Nothing is known on the intraspecific genetic variation in *S. capensis*. In studying this variation, information could be gained on the drainage history of the CFR, since the evolutionary history of the genetic variation observed today in a freshwater fish such as *S. capensis* is largely a reflection of the history of the river systems it occurs in. *Sandelia capensis* was believed to be a model species for our study for several reasons. The first is that *S. capensis* is one of the most widely distributed species in the CFR: it occurs naturally in the majority of Cape coastal rivers, from the Coega River (Algoa bay) on the south coast to the Verlorenvlei and Langvlei Rivers on the west coast (Skelton 1993). Another reason is that the distribution of lineages in *S. capensis* would be largely a reflection of its historical natural distribution, since it is not stocked in rivers or dams and movement by humans is therefore believed to be limited. There are, however, known cases of translocations, such as its introduction into the Clanwilliam Olifants system (Skelton 1993) and Eerste River (Impson, personal communication), as well as a translocation from the Keisers River to Hermanus according to a local farmer. A third important reason for studying the phylogeography of *S. capensis* is that it can be compared to other endemic fish of the CFR (Pseudobarbus and Galaxias) to further our

understanding of the aquatic biodiversity of this region. Finally, *S. capensis* seems to be able to tolerate a wide range of habitat conditions that might have contributed to its dispersal ability.

One of the aims of this project was to learn more about the drainage history of the CFR: what role river capture and/or sea level changes played in the evolution of *S. capensis* in different areas of its distribution. *Sandelia capensis* is also of conservation importance, since certain populations are declining as a result of many different threats (see Chapter 1). There are also other endemic fish species that co-exist with *S. capensis* that are endangered (for example Redfins and Cape galaxias). Another aim therefore, was to identify unique lineages in order to prioritize areas of importance for conservation action.

2.3 Materials and methods

Sample collection

Historical records from the Albany Museum and South African Institute for Aquatic Biodiversity (SAIAB) (both Grahamstown) of known *S. capensis* populations were used to identify priority localities for sampling and subsequent field surveys were undertaken with the aim of collecting representative samples from populations throughout the natural distribution of *S. capensis* (Figure 2, Appendix I). A total of 82 individuals from 21 of the sampled populations were included in the analysis (Figure 3, Table 1). The number of specimens analysed per population ranged from 1–6 (Table 1). Whole fish specimens were collected using seine and hand nets. In cases where it was difficult to find fish, fin clips were taken and the individuals were released. Fish and fin clips for genetic analysis were stored in 96% ethanol, while voucher specimens were stored in 10% formalin and deposited in the SAIAB fish collection. Not all populations were collected in this manner: specimens from the Boesmans, Verlorenvlei and Wolwekloof Rivers were collected, put on ice and sent to our laboratory by Mr. Dean Impson (Western Cape Nature Conservation Board). Immediately upon arrival, gill tissue was removed (see below) and individual specimens stored in sealed plastic bags at -20°C.

DNA extraction

Gills or fin clips were dried on paper towels, washed with ddH_2O, and finely chopped on petri dishes using sterile scalpel blades. Tissue was digested in a waterbath at 55°C overnight with extraction buffer (0.05 M Tris-HCl, 0.5 M EDTA·Na_2, 1.0 M NaCl, 10% SDS) and 0.5 mg of Proteinase K (Roche Diagnostics). Digestion of RNA was done by adding 0.06 mg RNase A (Roche Diagnostics) to each sample and placing them in a 37°C waterbath for 60 minutes. Total genomic DNA was extracted twice with phenol and then with a 24:1 solution of chloroform:isoamyl alcohol (Sambrook & Russell 2001). Precipitation of DNA was done overnight in a -20°C freezer by adding 0.1 volumes of sodium acetate (3M) and 2 volumes of ice cold 96% ethanol. Pellets were washed with 70% ethanol and then resuspended in SABAX (Adcock Ingram) or UHQ water. Extracted DNA was run on 0.8% agarose gels and visualised under UV-light.

PCR amplification and sequencing

A ±1 500 bp fragment was amplified using the polymerase chain reaction (PCR; Saiki *et al*. 1988). The fragment consisted of a large part of the cytochrome b gene, the threonine and proline tRNA genes and the 5' end of the control region. Primers used were a conserved vertebrate primer, L14841 (shortened primer of Kocher *et al*. 1989; numbered according to the human sequence published by Anderson *et al*. 1981), and a custom-made internal control region primer, H346 (Table 2).

A single PCR reaction (50µl volume) was set up as follows: 1.5mM $MgCl_2$, 1× reaction buffer, 0.2mM of each deoxy nucleotide triphosphate (Promega), 1.5 units Super-therm® Taq DNA polymerase (Southern Cross Biotechnology) and 50 pmol of each primer. Amplification was performed in a GeneAmp® PCR System 9700 (PE Applied Biosystems) with the following amplification profile: Initial denaturation at 94°C for 2 min, followed by 35 cycles of amplification: 30 sec denaturation at 93°C, 30 sec annealing at 50-54°C and 45 sec primer extension at 72°C. The reaction was completed

Table 1 Locality information and mtDNA haplotype distribution of *Sandelia capensis* populations that were analysed in the present study.

Population (River)	River System	GPS co-ordinates	N*	Haplotypes** CR and CB	CR	CB
Langvlei	Langvlei	32° 12' 30" S 18° 25' 30" E	3	Cb3(3)	D1(3)	C1(3)
Verlorenvlei	Verlorenvlei	32° 30' 10" S 18° 35' 50" E	5	Cb1(1), Cb2(1), Cb4(1)	D3(1), D2(3), D4(1)	C1(1), C2(1), C3(1)
Boesmans	Berg	32° 46' 30" S 18° 35' 00" E	5	Cb5(1)	D5(1), D6(2)	C4(3)
Diep	Diep	33° 31' 07" S 18° 40' 30" E	2	Cb6(1)	D7(2)	C5(1)
Palmiet	Palmiet	34° 04' 30" S 19° 02' 30" E	2	Cb7(1)	D8(2)	C6(1)
Heuningnes	Heuningnes	34° 34' 20" S 19° 56' 30" E	2	Cb8(1)	D9(1), D10(1)	C7(1)
Wolwekloof	Breede	33° 33' 50" S 19° 07' 55" E	5	Cb9(2)	D11(3)	C8(4)
Witte	Breede	33° 34' 30" S 19° 08' 30" E	3	Cb9(2)	D11(2)	C8(3)
Kleinkruis	Goukou	33° 00' 52" S 21° 17' 24" E	5	Cb10(3), Cb11(1)	D12(1), D13(3), D14(1)	C9(3), C10(1)
Langtou	Gourits	33° 58' 30" S 21° 47' 20" E	5	Cb12(1), Cb13(2)	D15(2), D16(3)	C11(1), C12(2)
Bos	Gourits	33° 43' 50" S 21° 30' 22" E	3	Cb14(1)	D16(1), D17(2)	C11(1)
Hoekplaas	Gourits	33° 34' 00" S 22° 58' 00" E	3	Cb15(1)	D18(2)	C13(2)
Groot Brak	Groot Brak	33° 58' 38" S 22° 11' 38" E	4	Cb16(3)	D19(4)	C14(3)
Malgas	Malgas	33° 56' 14" S 22° 25' 19" E	3	Cb18(1)	D22(3)	C15(1)
Touws	Touws	33° 58' 40" S 22° 36' 20" E	5	Cb17(3)	D20(4), D21(1)	C14(3)
Karatara	Karatara	33° 58' 55" S 22° 50' 11" E	5	Cb19(3), Cb20(1), Cb21(1)	D22(1), D23(3), D24(1)	C16(5)
Keurbooms	Keurbooms	33° 56' 17" S 23° 22' 00" E	4	Cb22(3)	D25(4)	C17(3)
Gheis	Kabeljous	33° 56' 26" S 24° 48' 17" E	1	-	D26(1)	-
Braam	Gamtoos	33° 43' 58" S 23° 57' 40" E	5	Cb23(5)	D27(5)	C18(5)
Maitlands	Maitlands	33° 57' 30" S 25° 16' 30" E	6	Cb24(1), Cb25(5)	D28(1), D29(5)	C19(1), C20(5)
Baakens	Baakens	33° 57' 32" S 25° 30' 53" E	6	Cb26(1), Cb27(3)	D30(1), D31(4)	C21(2), C22(3)

*Total number of individuals analysed per population.

**CR, control region; CB, cytochrome b; CR and CB, control region and cytochrome b data combined. Control region and cytochrome b haplotypes are preceded by the letters D (D-loop) and C (Cytochrome b) respectively. Combined haplotypes are preceded by Cb. Haplotype frequencies (absolute numbers) are given in parentheses.

with a final extention at 72°C for 5 min. Positive and negative controls were included in all reactions. Polymerase chain reaction products were visualised on 1.5% agarose gels.

Table 2 Names, sequences and annealing positions of primers used for PCR and sequencing. Primer L14841 is a conserved vertebrate primer (shortened primer of Kocher et al., 1989); all other primers used were custom-made.

Primer name	Primer sequence	Position**	Annealing region
L14841	5' CCAACATCTCAGCATGATGAAA 3'	15 366 – 15 387	Cytochrome b
H16091	5' GTATCATTCTGGTTTGATGTG 3'	16 091 – 16 111	Cytochrome b
L16547	5' TTACCCCTAACTCCCAAAGC 3'	16 528 – 16 547	TRNApro
H346	5'AGGAACCARATGCCAGKAATA 3'*	346 – 366	Control region

*R = A or G and K = G or T.
**Position according to the published sequence of the carp, *Cyprinus carpio* (Chang *et al*. 1994).

Polymerase chain reaction products were purified using the High Pure™ PCR Product Purification Kit (Boehringer Mannheim). The protocol provided with the kit was followed, except that 300µl of binding buffer was added to ±42µl of each PCR reaction. DNA was either purified directly or cut out of the gel (in cases of non-specific amplification) and eluted in either SABAX or UHQ water.

The 5' region of the cytochrome b gene was sequenced using primer L14841 and the custom-made primer H16091, based on *S. capensis* cytochrome b sequences (Table 2). The 5' end of the control region was sequenced using custom-made primers L16547 and H346, based on *Sandelia* tRNApro and control region sequences respectively (Table 2). Cytochrome b and control region were sequenced in both directions in order to double-check each base. The ABI Prism® BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems) was used for dye terminator (Sanger *et al*. 1977) cycle sequencing on a GeneAmp® PCR System 9700 (PE Applied Biosystems). Template DNA concentration used in each reaction ranged from 50-200ng. Initially half reactions in a total volume of 10µl were performed. The protocol followed was the one provided by the manufacturer, but with the following modifications: added additionally to each

reaction mixture was 2µl of 5× dilution buffer and 3.2 pmol primer was used (as for a full reaction). Subsequently quarter reactions were used, omitting dilution buffer.

The sodium acetate/EtOH method of DNA precipitation (PE Applied Biosystems) was followed: 10µl ddH_2O and 50µl 98% EtOH were added to each reaction, left at room temperature for 10min and centrifuged at 13 000rpm for 15-30min. The resulting DNA pellet was washed with 100µl 70% EtOH, spun down for 10min and dried on a heating block. Automated sequencing was carried out on ABI Prism 377 or 3100 DNA Sequencers (PE Applied Biosystems).

Sequence analysis

Light and heavy strand sequences were aligned and checked by eye in the Sequence Navigator program (PE Applied Biosystems). Multiple sequence alignments of consensus sequences were then performed using CLUSTAL X (Thompson *et al*. 1997). Unique haplotypes were identified using the TCS software package (Clement *et al*. 2000). Base frequencies as well as nucleotide (π) and haplotype (h) diversity values were estimated in Arlequin version 2.000 (Schneider *et al.* 2000). Haplotype diversity, h, is a measure of the frequencies and numbers of haplotypes among individuals, and varies between 0 and 1.0 (Nei & Tajima 1980). Haplotype diversity is the probability of two randomly drawn sequences being different in state. Nucleotide diversity (π) is the average weighted sequence divergence between haplotypes (Nei & Li 1979), and varies from 0% (no divergence) to over 10% (very high divergences).

Phylogenetic analyses

Separate analysis of data sets

Phylogenetic analyses were performed on control region and cytochrome b data sets separately, using PAUP version 4.0b (Swofford 1998). Evolutionary trees were constructed using distance, and parsimony tree building methods. In addition, maximum likelihood was performed on control region data. *Sandelia bainsii* could not be used as an

outgroup for *S. capensis*, since it was evolutionarily too divergent (see Chapter 3), even though they are believed to be sister species. Therefore, no outgroup was used for *S. capensis* and all evolutionary trees were constructed using midpoint rooting. Only unique haplotypes were used for tree construction. Models of DNA substitution that best fit the data were estimated using MODELTEST version 3.06 (Posada & Crandall 1998). The likelihood ratio test in MODELTEST identified the best-fit model of sequence evolution for control region to be HKY85 (Hasegawa *et al.* 1985) with mutation rates varying from site to site according to a gamma distribution (gamma shape parameter (α) estimated to be 0.1949). The model of nucleotide substitution that best fit the cytochrome b data set was estimated to be Tamura-Nei (Tamura & Nei 1993) with a gamma shape parameter of 0.1092. The proportion of invariable nucleotide sites was estimated in maximum likelihood to be 0.7 for control region data. These parameter values were then used to construct neighbour-joining (Saitou & Nei 1987) and maximum likelihood trees (Felsenstein 1981). Both maximum parsimony and maximum likelihood analyses used heuristic searches. Support for nodes were estimated using bootstrap (Felsenstein 1985) with 1000 replicates (neighbour-joining and parsimony analyses) or 100 replicates (maximum likelihood analysis).

Combining cytochrome b and control region data

The data sets of the slower evolving cytochrome b and the faster evolving control region were combined (total of 983 characters) in an attempt to provide more resolution at different parts of the phylogenetic tree. The partition homogeneity test in PAUP version 4.0b (Swofford 1998) was used to test for evidence of incongruence (Farris *et al.* 1995) between the two mtDNA regions. The test was performed on sequence data of twenty-seven individuals (unique haplotypes). Since no cytochrome b data were generated for the Kabeljous population, this population was excluded from the analysis. One thousand homogeneity replicates were specified and the heuristic search option was used with no branch-swapping. The most parsimonious tree constructed was 111 steps, and with P = 0.33 the null hypothesis of homogeneity could not be rejected (Farris *et al.* 1995). Control region and cytochrome b data were therefore combined for parsimony analysis.

The heuristic search option was chosen and a bootstrap analysis was performed with 1000 replicates. In addition, cladograms were constructed for each of the two data sets separately to check for any possible conflict.

Molecular clock analysis

To determine whether a molecular clock could be used for estimation of coalescence times, rate heterogeneity was tested with a likelihood ratio test in PAUP. Likelihood settings were those estimated by MODELTEST. We tested for rate heterogeneity by performing analyses both with and without the enforcement of a molecular clock. The difference between the two log likelihood scores obtained were then multiplied by two. This test statistic is then χ^2 distributed with n-2 degrees of freedom, where n is the number of haplotypes in the sample.

Analysis of molecular variance

Molecular variation among *S. capensis* populations was quantified using the analysis of molecular variance model (AMOVA; Excoffier *et al.* 1992). Variance components and Φ-statistics were calculated with 1000 permutations in the WINAMOVA program (Excoffier *et al.* 1992). Within a species there could be different levels of hierarchical subdivision: at the level of individuals (within populations), at the between-population level and between groups of populations. The Φ-statistics are a reflection of the correlation of haplotypic diversity at these different levels (Excoffier *et al.* 1992). The meaning of the different Φ-statistics is as follows: Φ_{ST} is the correlation of randomly drawn haplotypes within populations relative to that of randomly drawn pairs of haplotypes from the whole species, Φ_{CT} is the correlation of randomly drawn haplotypes within a group of populations relative to that of randomly drawn pairs of haplotypes from the whole species and Φ_{SC} is the correlation of randomly drawn haplotypes within populations relative to that of randomly drawn pairs of haplotypes from the region (Excoffier *et al.* 1992). The input distance matrix used for analysis consisted of pairwise

absolute number of differences for control region haplotypes. Similar results were obtained when an HKY85 distance matrix as input was used.

Mismatch distribution analysis

Mismatch analyses were also performed in Arlequin, in which 1000 bootstrap replicates were specified and 0% missing data allowed. A mismatch distribution is the frequency distribution of the observed number of pairwise genetic differences of DNA sequences (Rogers & Harpending 1992, Rogers 1995). The shape of the mismatch distribution provides information on the past demographic history of a population: whether the population has undergone a sudden ancient expansion, grown exponentially or whether its size has remained constant for a long time (Harpending 1994). Populations that have undergone ancient expansions or continued exponential growth, generate a mismatch distribution that is a smooth curve and has a single peak (a unimodal distribution). The position of this peak reflects the expansion time in units of mutational time. Mismatch distributions with many peaks are observed in populations that have not undergone rapid expansion (Rogers & Harpending 1992). In these populations a theoretical curve free of waves is expected, and there is very little agreement between data and theory (Rogers & Harpending 1992). Alternatively, multimodal distributions of pairwise differences can be generated by populations that are fragmented or geographically subdivided and different peaks are due to the mixture of different populations/lineages (Marjoram & Donnely 1994). The model of sudden expansion (Rogers & Harpending 1992) is a simple one: an initial population with a female effective population size of $N = N_0$ rapidly grows to a new size of $N = N_1$ at which it then remains, this sudden expansion is observed X units of mutational time later. The following three parameters (in units of mutational time) of population history (demographic expansion) were estimated: $\tau = 2\mu t$, $\theta_0 = 2N_0\mu$ and $\theta_1 = 2N_1\mu$, where τ is the time to the expansion, μ is the per generation probability of a mutation anywhere in the sequence, t is time of the expansion in generations, θ_0 and θ_1 is the population size before and after the expansion respectively (Harpending 1994, Schneider & Excoffier 1999). I also estimated the raggedness statistic, r, defined by Harpending (1994), which quantifies the smoothness of the observed mismatch

distribution. A population that went through an expansion will usually generate a mismatch distribution that is smooth (Harpending 1994). Populations that have remained constant in size and do not support the sudden expansion model, generate mismatch distributions with very ragged peaks (Rogers & Harpending 1992). A test of goodness-of-fit between the observed and theoretical mismatch distribution was done (Rogers 1995). In this test the parameters are declared compatible if the goodness-of-fit statistic falls within a 95% confidence interval obtained by simulation.

Neutrality tests were performed in Arlequin: Tajima's *D* test (Tajima 1989) and Fu and Li's *F** test (Fu & Li 1993). The two tests were performed with 1000 simulated samples and no missing data was allowed. Confidence limits for *D* followed Tajima (1989). If the *D* value falls within the confidence limit the value is not significantly different from 0, and the neutral mutation hypothesis can explain the observed DNA polymorphism (Tajima 1989). If *D* does not fall within this limit, the neutral mutation hypothesis is rejected ($P<0.05$). Significance of Fu and Li's *F* statistic followed Fu and Li (1993). Tajima's *D* values can also be interpreted together with mismatch distribution results, since negative *D* values are predicted by a model of sudden expansion (Aris-Brosou & Excoffier 1996, Bertorelle & Slatkin 1995, Guinand *et al.* 2001). Alternatively, positive values of the test statistics tend to be produced with population subdivision, an effect which becomes greater the more ancient the division (Simonsen *et al.* 1995).

Although it is possible to estimate the timing of an expansion by solving the formula: $t = \tau/2\mu$ (see first paragraph of this section), I decided against this since the generation time of *S. capensis* is uncertain (possibly 2 years, J. A. Cambray, personal communication).

Nested clade analysis

To separate population structure from population history, a cladistic analysis of the geographical distribution of *S. capensis* haplotypes was performed (Templeton *et al.* 1995). Nested cladistic analysis (NCA) can discriminate between phylogeographical associations due to ongoing restricted gene flow and those due to historical range expansion, fragmentation or colonization (Templeton *et al.* 1995). The first step in the

analysis was to estimate haplotype trees (gene genealogies) using the TCS software package (Clement *et al.* 2000). Haplotype trees were constructed for both cytochrome b and control region data sets. In the TCS program, outgroup probabilities are assigned to each haplotype, based on the correlation between haplotype frequency and haplotype age (Castelloe & Templeton 1994). Since the number of mutational differences between the majority of control region haplotypes were too great for them to be connected in the TCS program, only the gene genealogy estimated for cytochrome b haplotypes was used for NCA. All missing data was excluded, resulting in the use of 448 characters of cytochrome b for cladogram estimation. The next step was to convert the haplotype tree into a nested series of clades. This was done by pruning the tree starting at its tips and uniting all haplotypes that were separated by a single mutational step. Moving inwards, haplotypes (0-step clades) were nested into 1-step clades. This procedure was then repeated by grouping 1-step clades (connected by one mutational step) into 2-step clades. This nesting procedure was repeated until the entire haplotype tree was nested within a single clade. The estimated cladogram was converted into a nesting structure following nesting rules proposed by both Templeton *et al*. (1987) and Templeton & Sing (1993) as well as Cunningham (personal communication). The reason for following these alternative rules was to see how this would influence the resulting inferences from the key. Following Cunningham's nesting design, the haplotype tree was always pruned towards the root (which was the haplotype given the highest outgroup weight). With this method, loops of ambiguity were dealt with by breaking certain branches to give a single most parsimonious cladogram. In the design of Templeton *et al*. (1987) and Templeton & Sing (1993) nests are formed both by pruning from tips and also by expanding outwards from ungrouped internal clades. In this design nesting was performed over regions consisting of loops (Templeton & Sing 1993). Since there was uncertainty regarding two alternative nesting designs over loops, both designs were used for analysis and will be referred to as nesting designs A and B.

The nested cladistic analyses were performed using a computer program written in GWbasic (Cunningham, personal communication) and the software package GeoDis (Posada *et al.* 2000). The latter program was used on the nesting structure following

Templeton *et al*. (1987) and Templeton & Sing (1993). In these programs, latitude and longitude values were used to calculate two geographical distance measures: clade- (D_c) and nested clade distances (D_n). The statistical significance of these two distances was also determined. The clade distance is indicative of how widespread individuals within a particular clade are geographically. The nested clade distance is a measure of the distance of individuals in a particular clade from all individuals within the nesting clade (Templeton *et al.* 1995). Additionally, average interior minus average tip clade distances, $(I\text{-}T)_c$ and nested clade distances, $(I\text{-}T)_n$, were calculated, and their statistical significance assessed. These results were interpreted using the most recent inference key provided by Templeton (http://inbio.byu.edu/Faculty/kac/crandall_lab/geodis.htm).

2.4 Results

Sequence variation

Levels of variation in the cytochrome b gene were less than that observed for control region sequences from the same specimens. This was evident both from the smaller number of unique haplotypes identified and lower sequence divergence estimates (see later) obtained for cytochrome b.

Control region sequence variation

Approximately 390bp of control region sequence was obtained. Size differences among haplotypes were due to insertion/deletions (indels) ranging from 1-8 bp in size. A diagnosable size difference was observed between haplotypes from south and west coast populations. This difference was caused by a large indel (24bp) at the extreme 5' end of control region (Appendix II). In the ±390bp control region alignment, two regions (8bp and 11bp in size respectively) of ambiguity were excluded from all analyses, since they consisted of different sized gaps, making it difficult to code them for phylogenetic analyses. In the final 369bp alignment, 114 variable characters (70 transitions, 35 transversions and 9 indels) defined 31 unique haplotypes among 75 individuals from 21 populations (Table 1, Appendix II). This variation was observed at 89 nucleotide sites

(24.12%), of which 76 were parsimony informative and 16 were fixed differences between south and west coast populations. The following haplotypes were shared between populations: D11 from the Wolwekloof and Witte Rivers in the Breede system; D16 from the Langtou and Bos Rivers in the Gourits system; and D22 from Malgas and Karatara drainages (Table 1, Figure 3). There was an AT bias with base frequencies estimated to be 0.4029 (A), 0.1324 (C), 0.1273 (G) and 0.3374 (T). Haplotype diversity (*h*) and nucleotide diversity (π) values for within *S. capensis*, south coast- and west coast groups are indicated in Table 3. These measures of DNA polymorphism were among the highest estimated for freshwater fish (see review by Bernatchez & Wilson 1998). Typical haplotype and nucleotide diversity values found for other freshwater species ranged from 0.16–0.99 and 0.007–7.74 respectively (see review by Bernatchez & Wilson 1998). Of the species included in the review by Bernatchez & Wilson (1998), only seven displayed haplotype diversity values greater than 0.90. Only seven of these species had nucleotide diversity values greater than 3%.

Pairwise percentage sequence divergence estimates (HKY85 pairwise distances) among unique *S. capensis* haplotypes ranged from 0.27% to 15.38% (Appendix III, Table 4). Sequence divergence values were highest between west (Diep, Berg, Verlorenvlei and Langvlei populations) and south coast populations (populations from the Palmiet, Breede systems towards the Baakens), ranging from 9.85% to 15.38%, while values were 0.29-4.94% and 0.27-9.22% within west and south coast lineages respectively (Appendix III, Table 4). The Goukou and Maitlands drainages were each polymorphic for two divergent lineages: D12 and D14 vs. D13 haplotypes in the Goukou, and D28 vs. D29 in the Maitlands (Table 1, Figure 3).

The number of transitions and transversions between lineages are indicated in Table 5. Transition:transversion (TI:TV) values for the whole species were estimated to be 3.65:1 in MODELTEST and 2.56:1 in MacClade (Maddison & Maddison 1999). Values of 3.08:1 and 3.50:1 in MacClade (Maddison & Maddison 1999) were obtained for south and west coast populations respectively.

Cytochrome b sequence variation

Sequence for 614bp of cytochrome b was aligned for 60 specimens from 20 populations. Of the 81 variable nucleotide positions (13.19%), 62 were parsimony informative and 21 diagnostic between south- and west coast populations. In total, 85 variable characters (71 transitions and 14 transversions) delimited 22 unique haplotypes (Table 1, Appendix IV). Haplotypes were shared between Langvlei and Verlorenvlei drainages (haplotype C1), Wolwekloof and Witte (haplotype C8) populations from the Breede system, Bos and Langtou (haplotype C11) populations in the Gourits system and Groot Brak and Touws (haplotype C14) systems (Table 1). Seventy-one of the 81 variable sites (87.65%) were at the third codon position, nine (11.11%) in the first and one (1.23%) in the second position. Eleven differences (5.39%) were observed in the 204 codon amino acid alignment. Of these differences, one was fixed between the south- and west coast. Frequencies of the four bases were estimated to be 0.2565 (A), 0.2900 (C), 0.1571 (G) and 0.2964 (T). As for control region (see above), haplotype and nucleotide diversity estimates for the cytochrome b data set (Table 3) were also among the highest estimated for freshwater fish (Bernatchez & Wilson 1998).

The estimated sequence divergence among all unique *S. capensis* haplotypes ranged from 0.16% to 9.42% (Appendix III, Table 6). The cytochrome b divergences among lineages from south coast populations ranged from 0.16% to 3.1%, while values ranged from 0.22% to 2.24% among western lineages. Between south and west coast lineages divergence values were 6.69-9.42%. The divergent lineages identified within each of the Goukou and Maitlands drainages were the C9 and C10 haplotypes in the Goukou, and in the Maitlands C19 and C20 (Table 1, Appendix III).

The number of transitions and transversions between all haplotypes are shown in Table 7. The TI:TV ratio was estimated to be 5.37:1 for all haplotypes, and 6.82:1 and 3.57:1 for south- and west coast populations respectively.

Table 3 Nucleotide and haplotype diversity estimates for control region and cytochrome b data sets among *Sandelia capensis* haplotypes from 18 river systems. Values were estimated in the computer program Arlequin version 2.0 (Schneider *et al.* 2000).

	Control region		Cytochrome b	
Populations/Groups	h (SD)	%π (SD)	h (SD)	%π (SD)
Within *S. capensis*	0.970 (0.006)	6.5 (0.032)	0.950 (0.010)	3.1 (0.016)
Within west	0.897 (0.054)	2.5 (0.014)	0.800 (0.100)	0.9 (0.006)
Within south	0.960 (0.008)	4.4 (0.022)	0.936 (0.013)	1.4 (0.008)

Phylogenetic analyses

Control region

Parsimony analysis resulted in 75 equally most parsimonious trees (tree length = 142 steps; CI = 0.66; RI = 0.87). Distance, parsimony and maximum likelihood trees essentially had similar topologies. West coast populations formed a phylogenetic group distinct from south coast populations (Figure 4). The following phylogenetic groups from the major south coastal lineage (Figure 4) had high bootstrap support: Breede (D11) and Goukou (D13); Baakens (D30, D31) and Maitlands (D29); Bos (D16, D17) and Langtou (D15, D16); Heuningnes (D9, D10). The following lineages from the west coast group had good support: Langvlei (D1) and Verlorenvlei (D2, D3, D4); Langvlei, Verlorenvlei and Berg (D5, D6).

Cytochrome b

Parsimony analysis yielded 24 best trees (tree length = 90 steps; CI = 0.73; RI = 0.89). The phylogenetic distinction between south and west coast populations was also evident from cytochrome b gene trees (Figure 5). The cytochrome b data set gave less resolution than control region, with only three phylogenetic groups that had high support (Figure 5) in all trees: Baakens (C21, C22) and Maitlands (C20); Bos (C11) and Langtou (C11, C12); Gamtoos (C18) and Maitlands (C19).

Combined data

In the combined data set of 983 characters (369 for control region, 614 for cytochrome b), 137 were parsimony informative (75 for control region, 62 for cytochrome b). The heuristic search found eight most parsimonious trees (Figure 6; tree length = 232 steps; CI = 0.70; RI = 0.89). The most parsimonious trees constructed for the two mtDNA regions separately including only those individuals for which both fragments were sequenced, were 133 steps (12 trees, CI= 0.68, RI= 0.87) and 90 steps (24 trees) for control region and cytochrome b respectively (Figure 7). Control region and cytochrome b tree topologies were similar with very little conflict (Figure 7) and the following two clades were resolved for control region but unresolved for cytochrome b: the Baakens (D30, D31, C21, C22) and Maitlands (D29, C20) clade and the Breede (D11, C8), Goukou (D13, C9) and Palmiet (D8, C6) clade. The following phylogenetic groups within the major south coastal lineage were well supported (bootstrap ≥ 94%) in the combined analysis (Figure 6): Bos (Cb14) and Langtou (Cb12, Cb13) from the Gourits system; Gamtoos (Cb23) and Maitlands (Cb24); Baakens (Cb26, Cb27) and Maitlands (Cb25); Goukou (Cb10) and Breede (Cb9); Goukou, Breede and Palmiet (Cb7). The Groot Brak (Cb16), Touws (Cb17), Malgas (Cb18) and Karatara (Cb19, Cb20, Cb21) group (small coastal rivers) had less support (74%). Once again, there was good support (100%) for south and west coast as two distinct lineages. The following lineages were well supported within the major west coast group (Figure 6): Langvlei (Cb3) and Verlorenvlei (Cb1, Cb2, Cb4); Langvlei, Verlorenvlei and Berg (Cb5).

AMOVA

Genetic differentiation within and among S. capensis populations

Analysis of molecular variance results indicated substantial subdivision among *S. capensis* populations ($\Phi st = 0.911$, $P<0.001$): a very large fraction of the genetic variation was found among populations (91.07%), while only a small fraction existed within populations (8.93%). South coast populations were classified according to the four phylogenetic groups identified in the control region tree (Figure 4): Breede and Goukou; Baakens and Maitlands; Bos and Langtou; Karatara, Groot Brak, Touws and Malgas.

Population genetic structure among and within these regions was determined. In this hierarchical analysis, 69.08% of the total variance was due to variation among groups, 16.23% due to variation among populations within groups, and 14.69% due to variation within populations (ϕ_{CT} = 0.691, P<0.001; ϕ_{SC} = 0.525, P<0.001; ϕ_{ST} = 0.853, P<0.001). All south coast populations showed similar levels of structuring, with an among population variance component of 86.00% and within population variance component of 14.00% (ϕ_{ST} = 0.860, P<0.001). A significant amount of structure among the populations of the Gourits system was observed, with 91.76% of the variation among populations and only 8.24% of variation within populations (ϕ_{ST} = 0.918, P<0.001). A very large component of the total variance within the west coast group was due to among population variation (88.16%; ϕ_{ST} = 0.882, P<0.001), and a very small component due to within population variation (11.84%).

Mismatch analyses

Testing for an expansion

Results from the mismatch analyses (Table 8) together with Tajima's *D* and Fu and Li's *F* values are shown in Figures 8 and 9. Mismatch distributions that included all *S. capensis* populations were bimodal, the first and second peaks generated respectively by the smaller number of pairwise differences within southern and western groups, and the higher number of differences between the long isolated lineages of the two phylogenetic groups (Figures 8 and 9). Mismatch distributions observed for western and southern groups differed substantially (Figures 8 and 9). The mean number of pairwise differences observed within the southern group was almost twice as high as that within the western group for both mtDNA regions (Table 8). Mismatch distributions of the western group were multimodal and very ragged, with little agreement between the distribution of pairwise differences and an expected poisson distribution under a model of population expansion (Table 8, Figures 8 and 9). Multimodal distributions were also observed for the southern group, but these were much less ragged (Figures 8 and 9). The sudden expansion model was supported in all three groups (western group, southern group and within the species), except for control region data of the western group (Table 8).

Mismatch analyses in Arlequin specifying a DNA substitution model (with and without a gamma shape parameter), were also performed to see how this might influence the results. Most of these analyses caused the demographic expansion parameters to change very little or not at all (results not shown), and still supported the model of an ancient expansion in all three groups (again with the exclusion of western group control region data). However, when the K_2P model was specified (with and without the gamma shape parameter) for control region data of all *S. capensis* lineages (the HKY85 model is not given as an option in Arlequin), the demographic parameters changed dramatically: τ = 11.760, θ_0 = 10.613 and θ_1 = 51.660. However, when this substitution model was specified for the southern group, the demographic parameters remained similar: τ = 19.246, θ_0 = 0.816 and θ_1 = 36.223.

Tajima's D and Fu and Li's F values were all non-significant (with one exception), and we therefore failed to reject the neutral mutation hypothesis in these cases (Figures 8 and 9). The one exception was a significant positive F_s value obtained for control region data of the western group (Figures 8 and 9).

Nested clade analysis

The number of mutational steps separating south and west coast lineages were too great for TCS to connect them in the estimated cytochrome b cladogram. This resulted in the estimation of two disjoint haplotype networks, one comprising west coast lineages and the other south coast lineages. The southern lineage cladogram with alternative nesting designs A and B following the rules of Templeton *et al.* (Templeton *et al*. 1987; Templeton & Sing 1993) is shown in Figure 10. The cladogram of the southern group with the nesting design of Cunningham (personal communication) is shown in Figure 10. Estimated cladograms of the western lineage following the nesting designs of Templeton *et al.* (Templeton *et al*. 1987; Templeton & Sing 1993) and Cunningham (personal communication) are shown in Figures 12 and 13 respectively. Nested clades are designated by '*C-N*' where *C* represents the nesting level of the clade and *N* represents

Table 8 Control region and cytochrome b mismatch distribution results. Mismatch distribution analyses were performed on three specified groups: all *Sandelia capensis* populations, western- and southern populations. Mismatch observed means and demographic expansion parameters τ, θ_0 and θ_1 are expressed in units of mutational time. Shown also is $P(SSD_{obs})$, which is the probability that we might by chance observe a less good fit between the observed and simulated data, the raggedness index, r, and $P(Rag_{obs})$, the probability that we by chance might observe a higher value of r than the observed value under a model of sudden expansion.

Group	Mismatch observed mean*	*Tau*	θ_0	θ_1	$P(SSD_{obs})$	Raggedness index (r)	$P(Rag_{obs})$
Control region							
All populations	21.6 (190.8)	17.6	0.004	36.396	0.091	0.007	0.451
Southern group	15.6 (55.7)	20.1	0.000	47.322	0.554	0.006	0.435
Western group	8.6 (30.5)	15	0.002	18.562	0.031	0.177	0.042
Cytochrome b							
All populations	14.1 (171.7)	7.9	0.006	19.609	0.073	0.015	0.462
Southern group	7.1 (11.4)	8.7	0.000	33.516	0.245	0.017	0.232
Western group	4.3 (9.2)	6.5	0.002	8.123	0.102	0.245	0.071

*Mismatch observed variance given in parenthesis.

the clade number. Several most parsimonious cladograms were estimated for the southern lineage (Figure 10). The C14 haplotype from the Groot Brak and Touws Rivers was identified as ancestral (had the highest outgroup weight) to the southern clade while the Verlorenvlei and Langvlei lineage (C1 haplotype) was identified as ancestral to the western clade. All well-supported lineages identified in phylogenetic trees (estimated using parsimony, neighbour-joining and maximum-likelihood methods) were also identified as clades in all three nesting designs. It was the poorly supported lineages that resulted in the identification of different clades within the different nesting designs. An example of this was where the Keurbooms (C17) grouped into a two-step clade with one of the small south coastal rivers (the Malgas, C15) in nesting design B, while in design A the small coastal systems (C14, C15, C16) were in a two-step clade with the Hoekplaas (C13) from the Gourits system (Figure 10). In nesting design A (Figure 10a) the two oldest evolutionary clades (highest level clades) identified within the major south coast

lineage were a clade including all populations from the Goukou eastwards (which will be referred to as the Goukou-east clade) and a clade consisting of the Goukou, Breede, Palmiet and Heuningnes populations (which will be referred to as the Goukou-west clade).

South coast haplotypes showed significant geographical association following nesting methods of both Templeton and Cunningham (Figures 14, 15 and 16). Geographic association of haplotypes was also deduced for the western lineage following Templeton's nesting method (Figure 17). However, no structure was detected in this group following the method of Cunningham. Results obtained following the inference key of Templeton (http://inbio.byu.edu/Faculty/kac/crandall_lab/geodis.htm) are indicated in Tables 9 and 10.

Table 9 Interpretation of nested clade analysis results following the nesting procedure and inference key of Templeton *et al*. (http://inbio.byu.edu/Faculty/kac/crandall_lab/geodis.htm). Only clades with significant distance values are indicated.

Group	Clade	Populations in the nesting clade	Chain of inference	ΨInference
West coast	Haplotypes nested in total cladogram	Langvlei, Verlorevlei, Berg, Diep	1-2-11-12 No	Sequential colonization of adjacent river systems
South coast A*	§Haplotypes nested in clade 1-3	Baakens, Maitlands	1-2-3-5-6 Too few clades to determine concordance - Insufficient genetic resolution to discriminate between range expansion/colonization and restricted dispersal/gene-flow 7-8 No	Sampling design inadequate to discriminate between recurrent connections (short distance movements) versus unique historical connections between geographical distant river systems
	§Haplotypes nested in clade 1-6	Groot Brak, Malgas, Touws, Karatara	1-2 Tip/interior status cannot be determined	Inconclusive outcome
	§Haplotypes nested in clade 2-6	Palmiet, Breede, Goukou	1-2-11-12 No	Sequential colonization of adjacent river systems
	Haplotypes nested in clade 2-7	Groot Brak, Malgas, Touws, Karatara	1-2 Tip/interior status cannot be determined	Inconclusive outcome
	Haplotypes nested in clade 3-1	Goukou, Gourits, Groot Brak, Malgas, Touws, Karatara, Keurbooms, Gamtoos, Maitlands, Baakens	1-2-11-12-13-14 No	Sampling design inadequate to discriminate between sequential colonization of adjacent river systems and unique historical connections between geographical distant river systems
	Haplotypes nested in total cladogram	All populations	1-2 Tip/interior status cannot be determined	Inconclusive outcome
South coast B**	Haplotypes nested in clade 3-1	Groot Brak, Touws, Karatara, Gamtoos, Maitlands	1-2-11-12 No	Sequential colonization of adjacent river systems
	Haplotypes nested in clade 3-2	Goukou, Gourits, Malgas, Keurbooms, Maitlands, Baakens	1-2-3-5-6-7-8 No	Sampling design inadequate to discriminate between recurrent connections (short distance movements) versus connections between geographical distant river systems
	Haplotypes nested in total cladogram	All populations	1-2-11-12-13-14 No	Sampling design inadequate to discriminate between sequential colonization of adjacent river systems and unique historical connections between geographical distant river systems

* Results obtained following nesting method A (see text).

** Results obtained following nesting method B (see text).

§ Clades identified following both nesting methods (A and B), and which gave similar results. Ψ Interpretations from the inference key were rephrased in a way that is appropriate to *Sandelia capensis* (i. e. a stream fish).

Table 10 Nested clade analysis results following the nesting design of Cunningham (personal communication) and the inference key of Templeton *et al.* (http://inbio.byu.edu/Faculty/kac/crandall_lab/geodis.htm). Only groups and clades in which significant results were obtained are indicated in the table.

Group	Clade	Populations in the nesting clade	Chain of inference	ΨInference
South coast	Haplotypes nested in clade 1-3	Maitlands, Baakens	1-2-3-4 No	Recurrent connections among adjacent river systems
	Haplotypes nested in clade 2-8	Palmiet, Breede, Goukou	1-2-11-12-13-14 No	Sampling design inadequate to discriminate between sequential colonization of adjacent river systems and unique historical connections between geographical distant river systems
	Haplotypes nested in clade 4-2	Gourits (Langtou, Bos), Keurbooms	1-2-3-4-9-10 No	Geographical sampling scheme inadequate to discriminate between fragmentation and recurrent connections
	Haplotypes nested in clade 5-2	Gourits (Bos, Langtou), Keurbooms, Maitlands, Baakens	1-2-3-4-9-10 No	Geographical sampling scheme inadequate to discriminate between fragmentation and recurrent connections
	Haplotypes nested in clade 6-1	Goukou, Groot Brak, Touws, Malgas, Gamtoos, Maitlands	1-2-11-12 No	Sequential colonization of adjacent river systems
	Haplotypes nested in clade 7-2	Goukou, Gourits, Groot Brak, Touws, Malgas, Keurbooms, Gamtoos, Maitlands, Baakens	1-2-11-12 No	Sequential colonization of adjacent river systems
	Haplotypes nested in clade 7-1	Palmiet, Breede, Heuningnes, Goukou, Karatara	1-2-11-13-14 No	Sampling design inadequate to discriminate between sequential colonization of adjacent river systems and unique historical connections between geographical distant river systems
	Haplotypes nested in total cladogram	All populations	1-2-11-12 No	Sequential colonization of adjacent river systems

Ψ Interpretations from the inference key were rephrased in a way that is appropriate to *Sandelia capensis* (i. e. a stream fish).

2.5 Discussion

Genetic variation

In a review by Bernatchez & Wilson (1998) it has been shown that fish species from previously glaciated areas such as North America, Eurasia and Antarctica, show less intraspecific genetic variation than fishes occurring in areas that have not been previously glaciated (temperate regions). This study on *S. capensis* supports this: the intraspecific divergence and measures of DNA polymorphism found in *S. capensis* were much higher than those for northern species compared in the review by Bernatchez & Wilson (1998). Even in comparison to other temperate fishes, mtDNA variation in *S. capensis* was high. There are a few reasons why fish species from previously glaciated regions might have lower levels of genetic diversity in comparison to species from previously unglaciated regions (Bernatchez & Wilson 1998). First, glacial advances in northern areas would have caused major habitat destruction, resulting in the displacement (and destruction) of many fish populations (and subsequently, loss of genetic diversity). Populations would have been forced to survive in refugia during glacial periods, resulting in smaller population sizes (reduction in population sizes) and further loss of intraspecific genetic diversity. In addition, lineages within species from northern regions would appear to be evolutionarily younger with lower sequence divergence values, since the many glacial advances and retreats during the Pleistocene would have destroyed the original diversity that had been present in these species. Species from formerly non-glaciated regions would not have experienced the direct effects of glacial advances and retreats, and it is therefore expected that their intraspecific variation would be higher, as is the case with *S. capensis*.

Significant population structuring within *S. capensis* was supported by results from AMOVA. Most of the variation was due to between drainage variation, rather than within drainage variation. This was expected, since most populations were only from a single locality within drainages (with the exception of Breede and Gourits populations).

West and south coast lineages were genetically very distinct, with large divergence estimates between them: 9.85-15.38% for control region and 6.69-9.42% for cytochrome

b (Appendix III). These levels of intraspecific divergence between southern and western phylogenetic groups of *S. capensis* are among the highest reported for freshwater fish (Giuffra *et al.* 1994; see review by Bernatchez & Wilson 1998). Control region sequence divergence between south and west is up to twice that found between two distinct clades identified for the Berg River redfin (Bloomer & Impson 2000). The amount of genetic differentiation between these two lineages is comparable to what have been found between other freshwater fish species, for example among *Poecilia* (Ptacek & Breden 1998) and among salmonids (Bernatchez *et al.* 1992; Park *et al.* 2000).

Molecular clock

The molecular clock was used to provide an approximate time frame of divergence between lineages. It was not possible to calibrate the clock for *S. capensis*, since no fossil evidence exists for the Anabantidae. Consequently, clock calibrations reported in other fish studies were used as estimates. The average divergence rate for control region is considered to be 2% per my (Engelbrecht *et al.* 2000). The rate of control region variation in salmonids has been reported to be about 3% per my (Bernatchez & Danzmann 1993). Control region divergence rates for sturgeon have been reported to be as high as 8–10% per my (Brown *et al.* 1993). It was decided to use a tentative rate of 3–6% divergence per my for control region. Meyer (1994) suggested that the cytochrome b gene evolves at a rate 3–5 times slower than control region. In gasterosteid fishes cytochrome b divergence rate was estimated to be 2.8% per my (Orti *et al.* 1994). We applied both 2% and 4% divergence rates per my for cytochrome b. The values estimated for time since divergence between *S. capensis* lineages should be interpreted with caution, since they are only rough estimates.

Molecular clock estimations date the split between south and west to be of early-late Pliocene times, about 4-2 million years before present (BP). Estimated divergence times for all lineages within south- and west coast clades spanned the entire Pleistocene epoch (0.01–2 my BP). The most recent interconnections among CFR drainage basins were probably during the late Pleistocene (125 000-10 000 BP). Such recent connections might

for example have occurred between Langvlei and Verlorenvlei drainages in the west and the small coastal rivers between Mossel Bay and Knysna in the south, since these drainages have not only one of the lowest divergence values between them but they also share haplotypes (Figure 3). The most ancient divergence within the western group/lineage was estimated to be between the Berg and Diep drainages, dating back to middle Pleistocene times, ± 0.51–1.02 million years ago. In the south one of the oldest interconnections was estimated to be between the Heuningnes and Breede Rivers, a connection dating as far back as late Pliocene–early Pleistocene (estimated from control region data).

Drainage history

When considering the results from the traditional phylogeographic methods, NCA and mismatch distributions, it was concluded that a very complex evolutionary history (and therefore drainage history) was responsible for the genetic patterns observed within *S. capensis*. From molecular clock datings it is evident that changing climatic and environmental conditions in the last four million years (Pliocene–Pleistocene) might have played a major role in shaping this variation.

Unlike its South African counterparts (Redfin minnows and Cape Galaxias), *S. capensis* does not occur naturally in the Olifants (Clanwilliam) system, a drainage system just north of the species' natural distribution range on the west coast. Reasons for its absence are unknown. The Olifants and Berg River systems were connected during the early Tertiary (Hendey 1983). At that time it might have been possible for *S. capensis* to invade the Olifants from the Berg drainage. If *S. capensis* did indeed disperse to the Olifants during this time, a possible explanation for their absence from the system today would be that they had gone extinct. An alternative possibility is that *S. capensis* was unable to invade the Olifants during this time because it did not yet occur in the Berg system (and, possibly not in any of the west coastal rivers). If the latter is true, it is possible that *S. capensis* might have invaded Cape coastal rivers from the east (a scenario also suggested by Barnard 1943) and that, only after the separation of Olifants and Berg drainages, the species dispersed to west coast drainages. The invasion of the west from

the south (or vice versa: *S. capensis* could also have invaded south coastal systems from the west and subsequently gone extinct in the Olifants) could have been facilitated by the capture of the headwaters of the Breede system by the Little Berg, a tributary of the Berg River near Tulbach (Barnard 1943). Sea level fluctuations would have played no role in this invasion, since any river connections between south and west would have been impossible across the very steep and narrow continental shelf at the Cape peninsula.

From molecular clock evidence, fragmentation between west and south coastal lineages (very likely with the separation of Berg and Breede systems) is estimated to have taken place some time prior to the Pliocene or early Pliocene. The Hottentots Holland mountain range and the Cape peninsula seem to form the geological barriers between west and south coast lineages. The Little Berg managed to cut through these mountains at a certain stage in the past and have captured the Breede in the Tulbach valley (Barnard 1943).

Cytochrome b mismatch distributions and NCA (following Templeton's nesting procedure) supported a model of expansion (sequential colonization of adjacent river systems) within the west coast clade. These mismatch curves, however, were very ragged, which might suggest that fragmentation had played a major role in shaping the genetic variation found on the west coast. Conversely, west coast lineages showed no significant geographical association following Cunningham's nesting rules. It is possible that *S. capensis* extended its range on the west coast by first invading the Berg system and then dispersing to other west coast drainages (which could have been the expansion event). Although not supported by neutrality tests, mismatch distributions supported a model of expansion within the species and within the major south coastal lineage. For this lineage NCA could not discriminate between the sequential colonization of adjacent river systems and unique historical connections among geographically distant river systems (Templeton's design) as the historical processes responsible for the variation we see today. Following the rules of Cunningham the former was inferred for the south coast. The assumption made by mismatch analyses was that the observed genetic variation is the result of a single major ancient expansion. It seems unlikely, however, that the complex genetic patterns within *S. capensis* could be explained by a single

expansion event (or contiguous range expansion, as inferred by NCA). I contend that it is more likely the result of many smaller expansions during the different glacial periods of the Pliocene and Pleistocene. Sea level regressions and transgressions during the many Pleistocene glacial cycles would have resulted in the repeated connection and separation of river systems, especially within the south coastal lineage (see later). This would have made fish movements possible between different drainages at different times, probably resulting in many smaller connections (between geographically close systems), colonizations (connections between geographically distant systems) and short (recurrent connections) distance dispersal events. Overlaid upon these historical events are the processes of isolation or fragmentation (for example rivers sharing confluence becoming independent during interglacials), the signatures of which are evident from multimodal ragged mismatch distributions, high levels of structuring and high sequence divergence, haplotype and nucleotide diversity values.

During any of the major Pleistocene glacial periods the courses of south coastal CFR rivers extended up to 100 km seaward, and the Agulhas Bank would have been nearly 160 km wide south east of Cape Agulhas, 64 km wide off Cape St. Francis and 16 km wide to the west of the Cape Peninsula (Dingle & Rogers 1972). The exposed shoreline would have been rocky with small cliffs to the west of Cape Agulhas, and mostly sandy with some low cliffs in the east (Dingle & Rogers 1972) (Figure 1). The river valleys in the west would have been narrow, while the valleys in the east would probably have been much broader, marshy, sediment-choked and poorly drained with meandering rivers (Dingle & Rogers 1972). Dingle & Rogers (1972) postulated the extended courses of the Breede, Gourits, Keurbooms and Gamtoos Rivers during the Pleistocene low sea levels using bathymetric and sedimentological data. The Breede is postulated to have flowed in a south westerly direction while the Gourits, Keurbooms and Gamtoos Rivers extended their courses southwards (Figure 1). During these times the rivers flowed over a large area of low gradient, and they would have been wide. These three major drainage systems would have stayed independent from each other during these times.

This and other geological evidence, as well as the genetic evidence from this study were combined to start piecing together the evolutionary history of *S. capensis*. Results from NCA gave insufficient resolution mainly because of unsequenced locations and low sample sizes in general. However, useful information could still be gathered from NCA, since it not only helped to identify sampling gaps, but also made it clear that the evolutionary history of *S. capensis* is a complex one. The possible histories of certain groups of drainages, selected on the grounds of their geographical location and the phylogenetic groups identified in this study are discussed below.

The eastern drainages

The Gamtoos, Maitlands and Baakens are the three most easterly located rivers in the distribution range of *S. capensis* that were studied. The Maitlands was found to be polymorphic for lineages from two distinct clades (which will be referred to as the Gamtoos and Baakens clades/lineages). These two distinct Maitlands lineages diverged by values as high as 5.8% and 2.7% for control region and cytochrome b respectively, and Gamtoos and Baakens clades were separated by as many as 13 mutational steps in the haplotype tree (Figure 10).

It was not possible to determine which of the two clades are evolutionarily younger. Some evidence suggested that the drainage connection between Gamtoos and Maitlands is older: this clade was at a higher level in the cladogram, and it could only be identified as a clade in the slower evolving cytochrome b evolutionary tree. However, if the missing haplotypes between Gamtoos and Maitlands lineages were not sampled (and therefore not extinct) it might be younger than indicated by the cladogram. This possibility seems to be supported by molecular clock estimations which dates the split between Baakens and Maitlands lineages to be of roughly the same age as that between Gamtoos and Maitlands lineages (middle-late Pleistocene age or even older). The Gamtoos is a large, older drainage, while the Maitlands and Baakens are much smaller and believed to be much younger, their origin thought to date back to the late Tertiary, Pleistocene or even as recent as the Holocene (Skelton 1980b). The most likely scenario seems to be one in which the smaller Maitlands and Baakens drainages were colonized by fish from the

Gamtoos sometime during the Pleistocene (which might also have been the time of their origin). This is supported by the fact that the Gamtoos was ancestral to the Maitlands within the haplotype tree. The Baakens was also identified as ancestral to the Maitlands in the cladogram. These findings suggest two possible scenarios for the history of these three river systems. 1) It is possible that the Gamtoos and Maitlands lineages are present in all three river systems, but that only certain lineages/haplotypes were sampled (very high frequency of the one Maitlands lineage relative to the other). In this case an explanation might be that all three rivers had been connected at the same time during low sea level stands of the Pleistocene, with fish dispersing among them. 2) An alternative possibility is that the genetic pattern observed is a true reflection of the history and that the Maitlands contain divergent lineages from the Gamtoos and Baakens clades. This would imply that connections between the Maitlands and the other two systems took place at different times in the past and that the rivers were not connected simultaneously. In this case the Gamtoos would have shared a confluence with the Maitlands during the various sea level regressions of the Pleistocene, staying independent from the Baakens at these times. Fish movements between Baakens and Maitlands seem to have been possible only through river capture, since the Baakens flows into Algoa Bay whereas the Maitlands flows into St. Francis Bay (Figure 3).

The small south coastal lineage

The small river systems between Mossel Bay and Knysna are believed to form a single distinct evolutionary lineage, although with very low bootstrap support in phylogenetic trees (Figures 4 and 6). This low support is believed to be a consequence of similar genetic distances between haplotypes within this lineage, and between this lineage and other lineages, such as the Keurbooms. Support for this evolutionary lineage are the very low genetic distances between haplotypes (0.3-1.1% sequence divergence for control region), shared haplotypes among drainages and single mutational steps between its haplotypes in the estimated haplotype tree (Figure 10). In one nesting design this lineage was also identified as an evolutionary clade.

Molecular clock estimates date connections between the small south coastal drainages back to middle-late Pleistocene times. Geographically, these systems are located very close to each other, and it is therefore highly likely that they would have shared confluences on the exposed low gradient continental shelf of the south coast during periods of low sea level, as recent as the last glacial period.

Haplotypes within the small south coastal rivers were identified as ancestral to all other south coast lineages in the haplotype tree. This was unexpected, since these drainages are small and therefore believed to be much younger than the larger systems of the CFR (Skelton 1980b). One would therefore expect ancestral lineages to persist in the older, larger systems that have existed for much longer. The following however, is a possible explanation for the genetic pattern observed in these drainages. In certain control region trees the Hoekplaas (Gourits system) and Keurbooms grouped together with the small coastal lineage, although with low support. In addition, the Hoekplaas formed a clade (higher level clade and therefore older) together with the small coastal lineage in nesting design A (Figure 10a). The Gourits is a very large, ancient drainage, while the Keurbooms is also a large, older system, although smaller than the Gourits. Geographically, these two older drainages are located closest to and on either side of the much younger small south coastal rivers. From this and the above genetic evidence it is possible that the smaller rivers were colonized by fish from the Gourits and Keurbooms, probably through river capture.

The Gourits

Two divergent lineages were identified in the Gourits: the Bos and Langtou (which will be referred to as the Langtou lineage) and the Hoekplaas. The two clades differed by as much as 3.1-3.4% and 1.7-1.8% sequence divergence for control region and cytochrome b respectively, and were separated by no less than 5 mutational steps in the cladogram (Figure 10). There was a lack of significant population structure among the two populations of the Langtou clade, with a larger fraction of the variation within populations. When considering all three populations, there exists a significant amount of

structuring in the Gourits, with a very large component of the genetic variation due to differentiation between the Langtou and Hoekplaas lineages (see Results).

The Gourits and Keurbooms formed a higher level (older) clade following Cunningham's nesting procedure (Figure 10). Following Templeton's key, no distinction could be made between fragmentation and isolation by distance for this clade because of inadequate geographical sampling (Table 10). Therefore, a relatively early connection must have existed between the Gourits and Keurbooms systems, likely through river capture (an event also proposed by Skelton 1980a).

Where two clades meet: the Goukou

The Goukou, as the Maitlands, was polymorphic for divergent lineages. These lineages were separated by at least 10 mutational steps in the estimated haplotype tree and diverged with as much as 6.7% and 1.4% sequence divergence for control region and cytochrome b respectively. Each of the two divergent Goukou lineages were evolutionarily more closely related to each of the two oldest clades (the Goukou-west and Goukou-east clades) that were identified within the southern lineage of *S. capensis* (see NCA Results; Figure 10a). The Goukou therefore is a place where the Goukou-west and Goukou-east clades of the southern lineage meet.

The molecular clock suggests that the connection between Breede and Goukou drainages (within the Goukou-west clade) is more recent than that between the Goukou and rivers from the Goukou-east clade (the Gourits). Dingle & Rogers (1972) postulated that during the Pleistocene low sea level stands, the Goukou shared a confluence with the Gourits, while the Breede is postulated to have flowed in a south westerly direction, its mouth just south east of Cape Agulhas. In this case fish dispersal between rivers of the Goukou-east clade and Goukou would have been made possible through river confluence, while river capture event(s) must have allowed fish movements between rivers in the Goukou-west clade and the Goukou. The Agulhas Arch therefore seems to have kept the courses of the Breede and Goukou/Gourits separate during Pleistocene low sea levels, and this resulted

in the divergence of a Goukou-west and Goukou-east clade within the southern lineage of *S. capensis*.

The Palmiet, Breede and Goukou was identified as an evolutionary clade following all three NCA nesting designs (Figure 10). Within control region and combined phylogenetic trees, Breede and Goukou were evolutionarily more closely related to each other than either were to the Palmiet (Figures 4 and 6). This implies that the connection between Breede and Palmiet is older than that between the Breede and Goukou. However, the molecular clock dates the connections between Breede and these two drainages to be of roughly the same age, early-middle Pleistocene (1.5-0.75 my BP for control region and 0.4-0.2 my BP for cytochrome b).

The evolutionary mechanisms that shaped the genetic patterns observed in the Breede, Goukou and Palmiet clade was identified by NCA to be contiguous range expansion (sequential colonization of adjacent river systems) following Templeton's nesting design. With the method of Cunningham, sampling was inadequate to discriminate between range expansion and long distance colonization (unique historical connection between geographically distant river systems). If *S. capensis* invaded Cape coastal systems from the east the likely explanation for this clade would be dispersal to the Breede through a connection between it and the Goukou followed by colonization of the Palmiet from the Breede. River capture seems to be the most likely explanation for fish dispersal between Breede and Palmiet and Breede and Goukou respectively, because of the following two reasons: (1) confluence between Breede and Palmiet during Pleistocene low sea levels would not have been possible, since the exposed continental shelf to the west of Cape Agulhas would have been very steep and narrow with rocky terrain and narrow valleys, (2) the Goukou is postulated to have shared a confluence with the Gourits, not the Breede (Dingle & Rogers 1972).

A unique system: the Heuningnes

In all three NCA designs the Heuningnes was within a clade of its own (at the two-step level), and had at least five mutational steps between it and the closest haplotype

(Breede) (Figure 10). Within phylogenetic trees it was identified as a unique lineage that would group occasionally, but with low support, with the Breede, Goukou and Palmiet lineages (to form the western clade within the major south coastal lineage). When the molecular clock was enforced using control region data, the Breede-Heuningnes connection was estimated to be of late Pliocene–early Pleistocene age. In contrast however, cytochrome b data estimated the time to be middle Pleistocene. It is very likely that during Pleistocene glacials the Heuningnes flowed into the Breede as the latter extended its course in a south westerly direction with its mouth just south east of Cape Agulhas (Dingle & Rogers 1972). The genetic distinctness of the Heuningnes might seem to be inconsistent with such a connection, since the last confluence would have been as recent as the last interglacial, about 18 000 years BP. However, confluence between two systems does not necessarily imply dispersal among them.

The west coast lineage

West coast populations displayed high levels of structuring, with most of the genetic variation estimated to be between rather than within populations (see Results). These results together with ragged multimodal mismatch distributions suggest a high degree of isolation among the populations on the west coast (see Results). Three higher level clades were identified within the western lineage following Templeton's nesting structure (Figure 12): Langvlei and Verlorenvlei; Berg; and Diep. A minimum of seven mutational steps in the estimated haplotype tree separated the Diep from the other west coast lineages (Figures 12 and 13). The genetic distinction between Diep and other western lineages is also evident from high sequence divergence values (4.0-4.9% for control region and 2.1-2.2% for cytochrome b). From this evidence the Diep is clearly a very distinct and unique lineage.

From molecular datings it is evident that the connection between Berg and Diep is much older (early Pleistocene) than that between Berg and Verlorenvlei/Langvlei systems (late Pleistocene, 100 000 – 250 000 years BP). If *S. capensis* did indeed invade the west coast from the south, the very first arrivals would have been present in the Berg system. From here they would have dispersed to the Diep, probably via a river capture event since the

mouths of the Diep and Berg are too far apart for them to have ever shared a confluence in the Pleistocene or even Pliocene. Fish could have invaded rivers to the north of the Berg either by sharing a confluence during sea level regression or through river capture. However, the severing of Berg and Verlorenvlei/Langvlei systems is estimated to be much more recent. Langvlei and Verlorenvlei are the most close-by geographically, and these systems would have shared a confluence as recent as the last glacial period.

Relevance of the use of certain analyses in this study

Primary freshwater fishes such as *S. capensis* can only survive and disperse in freshwater and are therefore restricted to river systems (Skelton 1993) (this is in contrast with many terrestrial organisms that are able to disperse over land). At present the Cape coastal rivers are independent from each other, but in the past some of them shared confluence through sea level regression and some of them had been connected during river capture events. These are the only times that fish would have been able to move between different river systems. From this study, it became evident that certain analyses are more appropriate for use on organisms that are not restricted to freshwater habitats. Subsequently, results from certain analyses were interpreted with caution (mismatch distributions) while other analyses were adapted to suit *S. capensis* (NCA). With mismatch distributions, a possible ancient demographic expansion within a species can be identified. Although it is possible that *S. capensis* could have sequentially colonized adjacent river systems (an ancient demographic expansion event) when it invaded Cape coastal rivers, I think that most of the variation that we observe today is a consequence of many smaller expansions and colonization events, rather than a single, large expansion event. Nested clade analysis inferences of Templeton's key are not suitable for organisms that are restricted to freshwater habitats. Therefore, for the present study, the inference key was modified (in collaboration with Michael Cunningham) to be applicable to a freshwater fish such as *S. capensis* (see Tables 9 and 10).

Conservation implications

The main genetic findings of this study and subsequent conservation implications are summarized in Appendix V. West and south coast lineages of *S. capensis* have been historically isolated since at least the Pliocene, about 2-4 million years ago. These very divergent lineages would therefore, under Moritz' (1994b) definition of an evolutionarily significant unit (ESU), be regarded as two distinct ESUs. Following Moritz' (1994b) criteria for an ESU, nuclear DNA evidence is still needed in support of these two ESUs. The likelihood of finding nuclear allele frequency differences between south and west coast lineages is believed to be high since these lineages have been geographically separated for a long time. Also, it is unlikely that only males would have dispersed during later connections between the Berg and Breede, without any female dispersal taking place.

Following Crandall *et al*. (2000) west and south coast lineages are genetically non-exchangeable. It is not possible to say whether they are also ecologically non-exchangeable, since no comparative ecological studies have been done on the species. Therefore, following the criteria of Crandall *et al*. (2000) it would be impossible to say, at this stage, whether the two lineages should be treated as distinct populations or species. Crandall *et al*. (2000) stipulates three general principles that should govern management actions, and from these it is evident that south and west should be managed separately. The first principle is the preservation of adaptive diversity and evolutionary processes. Using Crandall's (2000) criteria we fail to reject the null hypothesis of exchangeability between south and west coast populations of *S. capensis*, but this does not necessarily mean that the null hypothesis is true, it simply means that there is a lack of relevant data such as morphological and ecological studies for this species. At this point in time I believe that the genetic evidence is strong enough support for conservation decisions in the light of managing these two lineages as separate conservation units, and no translocations should be allowed between them. The second principle is to preserve the natural network of genetic connections between populations. Since the major west and south coastal lineages are currently geographically isolated with no female gene flow between them since at least Pliocene times, this action of not allowing any translocations

between them would be a maintenance of the natural evolutionary processes in this species. Both the Diep and Heuningnes were identified as genetically distinct, and should be managed as such. Thirdly it is stated that conservation action might depend on the severity and nature of unnatural disturbances. In the case of *S. capensis* certain lineages should be prioritized for conservation action and should receive special conservation concern. One such lineage is the west coast ESU. This ESU should be regarded as threatened, since it is much smaller (both in number of populations and population sizes) than the southern ESU, and under threat, mainly from human activities and alien invasive fishes (Skelton 1993). The Diep River system is under severe threat from alien vegetation, the alien fish species *Tilapia sparmanii* and human activities such as total water extraction and cattle farming. The relatively small Verlorenvlei and Langvlei systems experience similar threats. Within the larger south coast lineage the genetically unique Heuningnes system is threatened by alien fish such as bass (*Micropterus* spp.), total water extraction and alien vegetation. Although the southern ESU and lineages identified within it seems to have larger population sizes and population numbers, it is also under pressure from human activities and aliens, and it is known that many populations are declining (see Introduction).

2.6 Conclusion

This study revealed the substantial amount of structuring that exists in *S. capensis* at the molecular level. The unique history of the CFR rivers is reflected by the genetic variation that is observed in the species today. River capture events (single and multiple) as well as confluence between geographically close drainages during periods of low sea level (multiple events: a result of the many glacial periods of the Pleistocene) had made fish dispersal between different drainages possible in the past. Off the Cape south coast the Agulhas bank and low gradient of the continental shelf seem to have played major roles in the genetic variation observed in this south coastal lineage. The low gradient of the shelf would have facilitated river confluence during sea level regressions, while the Agulhas bank seems to have kept the Breede and Goukou/Gourits separate, which very likely resulted in the origin of an eastern and western clade within the major south coastal lineage. Because of the steep, rocky continental shelf and narrow river valleys on the

west coast it is thought that sea level regression would have played a less important role, and that river capture events would have allowed fish to disperse. The two major lineages identified might have been kept separate by the Hottentots Holland mountains in conjunction with the steep continental slope on the west coast form geographical barriers between west (Lourens, Eerste Rivers) and south (Palmiet, Riviersonderend systems).

Each of the Heuningnes and Diep populations were found to be genetically very distinct from other lineages and should therefore be prioritised for conservation. These unique lineages should be managed as if they are ESU's until more evidence is gathered. The same action should be followed for all other well-supported lineages that were identified within *S. capensis* (Verlorenvlei/Langvlei, Berg, Breede/Goukou/Palmiet, Gourits, small coastal rivers, Gamtoos/Maitlands, Baakens/Maitlands). What contributes to the conservation importance of these drainages, are the severe pressures on them mainly from alien invasive fishes and agricultural practices such as water extraction. The major west coast clade should enjoy conservation priority as it is small (consists of few populations) and also under severe threat. Finally, translocations between the two major conservation units should not be allowed.

Chapter 3: A preliminary study on the phylogeography of *Sandelia bainsii* and phylogenetics of the Anabantidae

3.1 Abstract

The Eastern Cape Rocky, *Sandelia bainsii*, is classified as endangered and only a few small populations survive in coastal rivers of the Eastern Cape province, South Africa. A preliminary study was done to assess the genetic variation within this freshwater fish species. Two genetically distinct lineages were identified using mitochondrial DNA (mtDNA) control region sequences: a Great Fish/Kowie lineage and a Buffalo/Gulu lineage. The levels of genetic variation within these lineages seem to be low, which might support the hypothesis that populations of this species had gone through a bottleneck. We attempted to investigate the phylogenetic relationships of *S. bainsii* with *S. capensis* and other genera of the family Anabantidae using the mtDNA 16S rRNA gene. Deep divergences were found to exist between *S. bainsii* and *S. capensis*, and between the anabantid genera. The phylogenetic relationships among the anabantids could not be resolved, either because of possible saturation of the 16S rRNA gene or an ancient radiation within the family.

3.2 Introduction

The endangered *Sandelia bainsii* belongs to the family Anabantidae (Osteichthyes; Perciformes), in which there are three other genera: *Ctenopoma* and *Microctenopoma* (occuring further north in Africa) and *Anabas*, the only genus occurring in Asia (see Chapter 1). The phylogenetic relationships among the anabantid genera are unresolved (Norris 1994). Five monophyletic lineages (clades) have been identified in the family: *Anabas* (one clade), three clades within *Ctenopoma* and *Microctenopoma,* and *Sandelia* (one clade) (Elsen 1976; Norris & Teugels 1990). The Anabantidae appears to be the most primitive family in the suborder Anabantoidei (Liem 1963). On the grounds of

osteological evidence, Liem (1963) suggested that *Anabas* and *Ctenopoma* are derived from a common ancestral anabantid stock, a notion that is supported by their less specialised (generalised) morphology (Norris 1994). Liem (1963) suggested that this ancestral anabantid form invaded Africa from Asia where it differentiated into *Ctenopoma* in Africa and *Anabas* and other anabantoid families in Asia. He was also of the opinion that *Sandelia* had been derived from *Ctenopoma*, because of several specializations within *Sandelia*, such as its reduced suprabranchial organ. It was therefore hypothesized that *Ctenopoma* had spread from the favourable area of the African tropics to temperate South Africa, crossing the climatic barrier (to temperate regions) with difficulty. After crossing the barrier, *Ctenopoma* populations became isolated in the southern tip of Africa, giving rise to *Sandelia*. Skelton (1980), however, proposed an alternative hypothesis for the isolation and disjunct distribution of the anabantids and some cyprinids. He suggested that these groups originated in Africa, and their distribution is the result of continental drift (vicariance). Certain characters, for example geography and the presence of contact organs, support the monophyly of African anabantid lineages (Norris 1994). Subsequently, Norris (1994) recommended two subfamilies within the Anabantidae, the Anabantinae (Asian anabantids) and Ctenopominae (all African anabantids).

The Eastern Cape Rocky has a restricted distribution in coastal river systems of the Eastern Cape province (Figure 1) of South Africa (Cambray 1997a). The species occurs only in very short sections of the Nahoon, Buffalo, Keiskamma, Great Fish and Kowie systems (Skelton 1993). Recently, new populations have been discovered in the Gulu and Igoda drainages of the Eastern Cape (Cambray, personal communication). Many of these rivers undergo erratic flow as a result of prolonged droughts. Certain populations are on the verge of extinction, mainly because of threats such as water abstraction, damming, alien invasive fishes and plants and poor water quality (Cambray 1997a). Another major threat is water transfer schemes. An example of this is the underground Orange-Fish tunnel, which connects the Orange river with the Great Fish system. This tunnel has enabled sharptooth catfish, *Clarias gariepinus*, to invade the Great Fish river system, where it competes with the Eastern Cape Rocky (Cambray 1997b). Because of these

pressures, combined with small population sizes and a restricted geographical range, *S. bainsii* is listed as endangered (IUCN 2003).

Very few Cape Rockies are left in the Nahoon (although no *S. bainsii* could be found in this river during a survey in September 1996) and Buffalo rivers, where they compete with alien invasive fishes (Cambray 1997b). A small population (perhaps only 20 individuals) survives in the Bloukrans river near Grahamstown (Cambray, personal communication). Here the Blaauwkrantz Nature Reserve (managed by Makana Municipality) has been set aside especially for the conservation of the Eastern Cape Rocky (Cambray 1996). Another population in a tributary of the Keiskamma (the Tyume) is also very small and vulnerable to local extinction (Cambray 1996).

Immediate conservation action is needed to prevent extinction of the Eastern Cape Rocky. The conservation action proposed is restoration of the species' natural habitat and the use of farm dams for breeding programs and refuge areas (Cambray, personal communication). Knowledge on the genetic differentiation within the species and identification of conservation units would aid a great deal in management decisions regarding possible future breeding programs and the establishment of sanctuaries. A preliminary study was done on the levels of genetic variation within the Eastern Cape Rocky, using mitochondrial DNA (mtDNA) control region sequences. To investigate the phylogenetic relationships between *S. bainsii* and *S. capensis*, as well their relationships to the other anabantid genera, we used the mtDNA 16S rRNA gene as genetic marker. However, this study was preliminary, since sample sizes were very small: only one species from each of the anabantid genera (except *Sandelia*) were included in the study.

3.3 Materials and methods

Sample collection

Ten specimens from five anabantid species (all four genera were represented) were analysed for the phylogenetic study (Table 1). The five species studied were: *Anabas testudineus*, *Microctenopoma intermedium, Ctenopoma multispine, Sandelia bainsii* and *S. capensis* (Table 1). *Microctenopoma intermedium* and *C. multispine* tissue was

collected, stored in 96% ethanol and sent to us by Mr. Roger Bills (South African Institute for Aquatic Biology). *Anabas testudineus* samples were also provided by Mr. Roger Bills. *Sandelia capensis* DNA samples (one sample from each of four populations: Langvlei, Diep, Gourits and Baakens drainages) were included in the study (see Chapter 2). As outgroup for the Anabantidae we used a Pearl Gourami (*Trichogaster leeri*) that was obtained from the aquarium trade. *Sandelia bainsii* samples were collected by Dr. J. A. Cambray (Makana Biodiversity Centre, Albany Museum, Rhodes University, Grahamstown) using electro-shocking fishing gear (which is the most efficient sampling method in the species' rocky habitat). Since *S. bainsii* populations are threatened, fin clips were taken and fish returned to the river. Fin clips were stored in 96% ethanol and sent to our laboratory for analysis. Eleven specimens from five populations were included in the study (Figure 2, Table 2). The number of individuals analysed per population ranged from 1-3 (Table 2).

Table 1 Locality information and mtDNA 16S ribosomal RNA gene haplotype distribution of 10 individuals from five anabantid species. All four genera of the family Anabantidae were represented.

Species	River	River system	GPS co-ordinates	N*	Haplotypes
Anabas testudineus§	Malaysia, aquarium trade		-	1	A1
	Tutong, Brunei		-	1	A2
Ctenopoma multispine§	Kwazulu Natal		-	1	A3
	Sepopa, Botswana		18° 44' 39" E 22° 11' 47" S	1	A4
Microctenopoma intermedium	Lower Zambezi	Zambezi	18° 21' 39" S 35° 53' 57" E	1	A5
Sandelia bainsii	Koonap	Great Fish	32° 49' 33" S 26° 08' 18" E	1	A6
Sandelia capensis	Baakens	Baakens	33° 57' 32" S 25° 30' 53" E	1	A7
	Diep	Diep	33° 31' 07" S 18° 40' 30" E	1	A8
	Langtou	Gourits	33° 58' 30" S 21° 47' 20" E	1	A9
	Langvlei	Langvlei	32° 12' 30" S 18° 25' 30" E	1	A10

*Number of individuals analysed.
§No GPS co-ordinates available.

Table 2 Locality information and mtDNA control region haplotype distribution of *Sandelia bainsii* populations that were analysed.

Population number	Population (River)	River system	GPS	N*	Haplotypes**
1	Bloukrans	Kowie	33° 23' 25" S 26° 42' 26" E	2	B1(1), B2(1)
	Bloukrans	Kowie	33° 27' 28" S 26° 41' 32" E		
2	Koonap	Great Fish	32° 49' 33" S 26° 08' 18" E	3	B1(2§), B3(1)
3	Kat	Great Fish	32° 38' 54" S 26° 39' 52" E	1	B1(1§)
4	Yellowwoods	Buffalo	32° 44' 13" S 27° 31' 20" E	2	B4(1+1§)
	Yellowwoods	Buffalo	32° 43' 29" S 27° 29' 20" E		
5	Gulu	Gulu	32° 42' 38" S 27° 42' 38" E	3	B4(3)
Total number of individuals analysed			11		
Total number of haplotypes identified			4		

*Total number of individuals analysed per population.
**Control region haplotypes that were identified. Haplotype frequencies (absolute numbers) are given in parentheses.
§Haplotypes that were sequenced only in one direction.

DNA extraction

Total genomic DNA was extracted following the methods described in Chapter 2. Extracted DNA was run on 0.8% agarose gels and visualised under UV-light.

PCR amplification and sequencing

The PCR technique (Saiki *et al*. 1988) was used for amplification of a ±600 bp fragment close to the 3' end of the 16S ribosomal RNA gene of the five anabantid species. The universal primers 16Sar and 16Sbr (Table 3) from Palumbi *et al*. (1991) were used for amplification. A ±370 bp fragment of the 5' end of the mtDNA control region of *S. bainsii* was amplified using the same custom-made primers (L16547 and H346, Table 3) that were used to amplify the mtDNA control region of *S. capensis* (Chapter 2). Polymerase chain reaction were set up as in Chapter 2. I used 50pmol of control region primers and 0.25 pmol of each 16S rRNA primer. The same amplification profile as in Chapter 2 was used for both mtDNA regions, with annealing temperatures of 50-52°C for control region and 50°C for the 16S rRNA gene. Positive and negative controls were included in all reactions. Polymerase chain reaction products were visualised on 1.5% agarose gels.

Polymerase chain reaction products were purified (see Chapter 2) using the High Pure™ PCR Product Purification Kit (Boehringer Mannheim). Each of the mtDNA regions were sequenced with the same primers used for their amplification (Table 3). Except for three *S. bainsii* control region sequences (Table 1), sequencing was done in both directions for both genes in order to double-check each base. Cycle sequencing (see Chapter 2) was carried out on a GeneAmp® PCR System 9 700 (PE Applied Biosystems). Template DNA concentration used in each reaction ranged from 50-200ng. Quarter reactions using 3.2pmol primer (total reaction volume: 10µl) were performed. Precipitation of DNA was done following the sodium acetate/EtOH method (see Chapter 2). Automated sequencing was carried out on ABI Prism 377 and 3100 DNA Sequencers (PE Applied Biosystems).

Table 3 Names, sequences and annealing positions of primers used for PCR and sequencing. Custom-made primers were used for amplification of the control region of *Sandelia bainsii* (see Chapter 2). The 16S rRNA gene of the five anabantid species was amplified using the universal primers 16Sar and 16Sbr from Palumbi *et al*. (1991).

Primer name	Primer sequence	Position**	Annealing region
L16547	5' TTACCCCTAACTCCCAAAGC 3'	16 528 – 16 547	TRNApro
H346	5' AGGAACCARATGCCAGKAATA 3'*	346 – 366	Control region
16SAR	5' CGCCTGTTTATCAAAAACAT 3'	2 919-2 938	16S ribosomal RNA
16SBR	5' CCGGTCTGAACTCAGATCACGT 3'	3 520 – 3541	16S ribosomal RNA

*R = A or G and K = G or T.
**Position according to the published sequence of the carp, *Cyprinus carpio* (Chang *et al*. 1994).

Phylogenetic analyses

Light and heavy strand sequences were aligned and manually checked in the Sequence Navigator program (PE Applied Biosystems). Multiple sequence alignments of consensus sequences were done in the computer program CLUSTAL X (Thompson *et al*. 1997). Unique haplotypes were identified using the TCS software package (Clement *et al*. 2000).

Phylogenetic analyses were performed using the PAUP version 4.0b computer program (Swofford 1998). Evolutionary trees were constructed using distance, parsimony and maximum likelihood tree building methods for the 16S rRNA data set. Only distance and maximum likelihood methods could be used to construct the *S. bainsii* control region tree (there were no parsimony informative characters in this data set). For the five anabantid species, a gourami was specified as outgroup. Since *S. capensis* was evolutionarily too divergent from *S. bainsii,* no outgroup could be specified for this species and midpoint rooting was used to construct phylogenetic trees. Only unique haplotypes were used for construction of trees. Models of DNA substitution that best fit the data were estimated using the program MODELTEST version 3.06 (Posada & Crandall 1998). The model of nucleotide substitution that best fit the 16S rRNA data set was estimated to be Kimura's two parameter model (Kimura 1980) with mutation rates varying from site to site according to a gamma distribution (α = 0.2085). The likelihood ratio test in MODELTEST identified the best-fit model of sequence evolution for the control region of *S. bainsii* to be HKY85 (Hasegawa *et al.* 1985), with equal rates for all sites. Neighbour-joining (Saitou & Nei 1987) and maximum likelihood (Felsenstein 1981) trees were constructed using these parameter values. Parsimony and maximum likelihood analyses used heuristic and branch-and-bound searches respectively. Support for nodes was estimated using the bootstrap criterion (Felsenstein 1985) with 1000 replicates (neighbour-joining and parsimony analyses) or 100 replicates (maximum likelihood analysis).

3.4 Results

Sequence variation

16S sequence variation

Approximately 486 bp of 16S rDNA sequence was obtained (Appendix VI). Size differences among haplotypes were due to insertion/deletions (indels). A total of 131 variable characters (74 transitions, 54 transversions and 3 indels) defined 10 unique haplotypes among the five anabantid species (Table 1, Appendix VI). This variation was

observed at 107 nucleotide sites (22.02%), of which 79 were parsimony informative. Base frequencies were estimated to be equal.

Pairwise percentage sequence divergence estimates (K2P pairwise distances) among anabantid genera (Table 4) ranged from 6.06% (between *Ctenopoma* and *Sandelia*) to 15.10% (between *Anabas* and *Microctenopoma*). The estimated sequence divergence values between the two *Sandelia* species were as high as that found between some genera, and ranged from 7.4–8% (Table 4). The outgroup diverged with the anabantids from 8.71% (with *Ctenopoma*) to 15.38% (with *Anabas*). The highest divergences were between Asian and African anabantids, ranging from 12.93% to 16.54% (Table 4). The lowest divergence values found between any two species were between *Ctenopoma multispine* and *S. capensis* (5.63-6.65% percentage divergence). The highest divergence between any two species ranged from 16.52-16.54% between *S. bainsii* and *A. testudinues*. The number of transitions and transversions between lineages are indicated in Table 5. A TI:TV ratio of 2.06 was estimated in MODELTEST.

S. bainsii control region variation

Control region sequences of about 339 bp in length were obtained for 11 individuals from five populations (Appendix VII). Four unique haplotypes were defined by a total of 21 variable characters (13 transitions, 7 transversions and 1 insertion/deletion). None of the 21 variable nucleotide sites (6.19%) contained any parsimony informative characters (Appendix VII). Nineteen (18 substitutions and 1 insertion/deletion) of the 21 variable characters were diagnostic between Great Fish/Kowie and Buffalo/Gulu populations (Appendix VII). Populations from the Buffalo and Gulu drainages had a single unique haplotype, B4 (Table 2). Three unique lineages (B1, B2, B3) were indentified within the Great Fish and Kowie systems (Table 2). The B1 haplotype was shared between Great Fish and Kowie systems, while the B2 and B3 haplotypes were unique to the Kowie and Great Fish drainages respectively (Table 2). Frequencies of the four bases were estimated to be 0.3877 (A), 0.1522 (C), 0.1533 (G) and 0.3068 (T) in MODELTEST.

Percentage sequence divergence values (HKY85 pairwise distances) ranged from 0.3-5.91% among the four lineages (Table 6). Divergence values were the highest (5.58-5.91%) between Great Fish/Kowie and Buffalo/Gulu populations. Among lineages from the Buffalo and Gulu; and the Great Fish and Kowie systems divergence values of 0% and 0.3-0.59% respectively were the lowest (Table 6). The pairwise number of transitions and transversions between all lineages are indicated in Table 6. The TI:TV ratio was estimated to be 1.94 in MODELTEST for all haplotypes.

Table 6 Pairwise HKY85 percentage genetic distances (below diagonal) and number of transitions\transversions (above diagonal) among four mitochondrial DNA control region haplotypes identified in *Sandelia bainsii*. Haplotype designations follow Table 2.

	B1	B2	B3	B4
B1		0\1	1\0	12\6
B2	0.3		1\1	12\7
B3	0.3	0.59		13\6
B4	5.58	5.9	5.91	

Phylogenetic analyses

The five anabantid species

The topologies of distance, parsimony and maximum likelihood phylogenetic trees were similar, therefore, only a distance tree is presented (Figure 3). Parsimony analysis resulted in two equally most parsimonious trees (tree length = 132 steps; CI = 0.8; RI = 0.8). The phylogenetic relationships between the four anabantid genera were unresolved (Figure 3). There was low bootstrap support (62%) on the neighbor-joining tree, and no support in the parsimony and maximum likelihood phylogenetic analysis for the two *Sandelia* species grouping together within a monophyletic clade (Figure 3). The two major phylogenetic groups (west coast and south coast) identified in *S. capensis* (Chapter 2) had high bootstrap support (=82%).

Sandelia bainsii

Distance and maximum likelihood trees had similar topologies. Two distinct lineages were identified, a Great Fish/Kowie lineage (haplotypes B1, B2, B3; Table 2) and a lineage (haplotype B4) from the Buffalo and Gulu drainages (Figure 4).

3.5 Discussion

Divergence values between African and Asian anabantid genera were as high as that found between them and the outgroup (Table 4). One possible explanation for this might be saturation of the 16S rRNA gene due to multiple substitutions. If this is the case, deeper phylogenetic relationships in the 16S rDNA tree would be misleading. An alternative explanation could be that outgroup choice was poor: it is believed that the Anabantoidei was derived from a common ancestral anabantid stock, with the Anabantidae as the most primitive of the Anabantoid families (Liem 1963, Norris 1994). An alternative explanation to saturation could be that the Anabantidae is an old family, and that the short branch lengths (Figure 3) deeper in the tree (at internal nodes) are indicative of a rapid ancient radiation within the family. The highest divergences estimated were between the Asian *Anabas* and the African anabantids (Table 4). This would be expected, since not only geographical, but also osteological data support such a deep divergence between Asian and African anabantids (Norris 1994). No inferences could be made about the phylogenetic relationships among the anabantid genera, since interior nodes had little support.

Sandelia bainsii and *S. capensis* are classified as sister taxa, within a single genus, *Sandelia* (Norris 1994). However, divergence values estimated between these two species approached those found between genera of anabantids (Table 4). In the phylogenetic tree, very low support (62% by distance, none by parsimony and maximum likelihood) was given to *Sandelia* as a phylogenetic group (monophyletic lineage). The large genetic difference in relatively conserved mtDNA genes is surprising and although the present study cannot reject the possibility of monophyly between the two *Sandelia* species, their relationships to other anabantids should be investigated further with more conservative markers, especially nuclear genes.

Two genetically distinct lineages were discovered within *S. bainsii*: a lineage (a single haplotype) from Buffalo and Gulu drainages (in the north), and another (consisting of three haplotypes) from the Great Fish and Kowie systems (in the south). The sequence divergence of 5.58-5.91% between the two lineages is not as high as that found between two distinct lineages within *S. capensis* (9.85-15.38%) (see Chapter 2), but it was similar to divergence values of 5.3-7.03% found between two distinct Berg River Redfin lineages (Bloomer & Impson 2000). This implies that drainages containing each of the lineages have been historically isolated, and have not been connected recently. If a molecular clock (see Chapter 2) is applied, the split between the two lineages is estimated to be of Pleistocene times (estimated to be 1-2 mya). It will be important to analyze individuals from the Keiskamma System, since it is located in an area between the above mentioned lineages.

Within the two major lineages identified, divergence values were very low, ranging from no divergence (a single haplotype in the Buffalo and Gulu rivers) to 0.59% (between the Great Fish and Kowie systems). Since this was only a preliminary study for assessing the genetic diversity within *S. bainsii*, sample sizes were small. Consequently, it is not possible to know for sure whether this low diversity found within each of the two major lineages is real. If low genetic diversity proves to be the case for *S. bainsii* (with the inclusion of more samples for future study), the idea that populations have gone through severe bottlenecks would be supported. This is a highly likely scenario, since most if not all *S. bainsii* populations are under severe threat and extremely small in size. One example of such a population is the Blaauwkrantz Nature Reserve population in which only approximately 20 mature individuals are left within a short section of the Bloukrans river which are now under pressure from invasive alien bass (*Micropterus salmoides*) (Cambray, personal communication).

Large drainages of the eastern Cape, such as the Great Fish, would have expanded naturally in the past, coming in contact with adjacent systems (Skelton 1980b), making it possible for fish to disperse among them. Such events, within each of the two major

lineages could have contributed to the low variation we observe within them today. In the past, the Bloukrans extended south east, toward the Kat, a tributary of the Great Fish river (Skelton 1980). The fact that Kowie and Great Fish populations share a haplotype support a historical connection between them.

The genetic variation (two highly divergent lineages) identified within *S. bainsii* should be conserved, it is recommended that individuals from the two different lineages are not intermixed. The most important conservation action must be the establishment of sanctuaries in several areas in order to protect all lineages. Without such sanctuaries *S. bainsii* might not survive. If there are to be a breeding program for this species, it should be for both lineages separately, this will maximise the genetic diversity within the species.

3.6 Conclusion

The phylogenetic relationships among the anabantid genera were unresolved using the mtDNA 16S rRNA gene as genetic marker. Alternatively, there could have been a rapid ancient radiation within the family. To resolve these relationships it would be necessary to use one or more nuclear markers (which are slower evolving), since the divergences between the anabantid genera seem to be very old and/or saturated. For a complete phylogenetic study of the Anabantidae, all species should be included. The reasons for including all members and good geographical sampling is that gaps might result in larger divergences being estimated. If possible, the two historically isolated lineages identified within the endangered Eastern Cape Rocky should be conserved separately and no translocations should be allowed between the drainages these two lineages occur in. Populations from the Keiskamma drainage (located between the two major lineages) should be analysed in order to identify the border between the two lineages. The levels of genetic variation within each of the major lineages seem to be low, which would support the hypothesis that *S. bainsii* populations had gone through a bottleneck. However, sample sizes should be increased in order to confirm this.

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature,* **290**, 457-465.

Aris-Brosou S, Excoffier L (1996) The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Molecular Biology and Evolution,* **13**, 494-504.

Avise JC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, **18**, 489-552.

Avise JC (2000) Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge, Massachusettes London, England.

Barnard KH (1943) Revision of the indigenous freshwater fishes of the S.W. Cape region. *Annals of the South African Museum*, **36**, Pt. 2, 101-262.

Bernatchez L, Guyomard R, Bonhomme F (1992) DNA sequence variation of the mitochondrial control region among geographically and morphologically remote European brown trout *Salmo trutta* populations. *Molecular Ecology,* **1**, 161-173.

Bernatchez L, Danzmann RG (1993) Congruence in control-region sequence and restriction-site variation in mitochondrial DNA of Brook charr (*Salvelinus fontinalis* Mitchill). *Molecular Biology and Evolution,* **10**, 1002-1014.

Bernatchez L, Wilson CC (1998) Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology,* **7**, 431-452.

Bertorelle G, Slatkin M (1995) The number of segregating sites in expanding human populations, with implications for estimates of demographic parameters. *Molecular Biology and Evolution,* **12**, 887-892.

Bloomer P, Impson ND (2000) Mitochondrial DNA differentiation in the critically endangered Berg River redfin (*Pseudobarbus burgi*). *Journal of Heredity,* **91**, 122-127.

Bowen BW (1999) Preserving genes, species, or ecosystems? Healing the fractured foundations of conservation policy. *Molecular Ecology*, **8**, S5-S10.

Brown JR, Beckenbach AT, Smith MJ (1993) Intraspecific DNA sequence variation of the mitochondrial control region of White Sturgeon (*Acipencer transmontanus*). *Molecular Biology and Evolution*, **10**, 326-341.

Cambray JA (1980) Contact organs in fishes of the genus *Sandelia* (Anabantidae). *South African Journal of Science*, **76**, 277-279.

Cambray JA (1990) The early ontogeny of the Cape Kurper, *Sandelia capensis* (Pisces: Anabantidae). *Annals of the Cape Provincial Museums (Natural History)*, **18**, 159-182.

Cambray JA (1996) Threatened fishes of the world: *Sandelia bainsii* Castelnau, 1861 (Anabantidae). *Environmental Biology of Fishes,* **45**, 150.

Cambray JA (1997a) The spawning behaviour of the endangered Eastern Cape rocky, *Sandelia bainsii* (Anabantidae), in South Africa. *Environmental Biology of Fishes,* **49**, 293-306.

Cambray JA (1997b) Captive breeding and sanctuaries for the endangered African anabantid *Sandelia bainsii*, the eastern cape rocky. *Aquarium Sciences and Conservation,* **1**, 159-168.

Cambray JA (2004) Spawning behaviour of *Sandelia capensis* (Anabantidae) (In Press).

Castelloe J, Templeton AR (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. *Molecular Phylogenetics and Evolution,* **3**, 102-113.

Chang YS, Huang FL, Lo TB (1994) The complete nucleotide sequence and gene organization of carp (Cyprinus carpio) mitochondrial genome. *Journal of Molecular Evolution,* **38**, 138–155.

Citation: IUCN 2003. *2003 IUCN Red List of Threatened Species*. <www.redlist.org>. Downloaded on **15 December 2003**.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology,* **9**, 1657-1659.

Crandall KA, Bininda-Edmunds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution*, **15**, 291-295.

Dingle RV, Rogers J (1972) Pleistocene palaeogeography of the Agulhas bank. *Transactions of the Royal Society of South Africa,* **40**, 155-165.

Elsen M (1976) La vessie gazeuse it l'organe labyrinthique des anabantidae. *Bulletin de Academie royaume de Belgique*, **62**, 49-79.

Engelbrecht CC, Freyhof J, Nolte A, Rassman K, Schliewen U, Tautz D (2000) Phylogeography of the bullhead *Cottus gobio* (Pisces: Teleostei: Cottidae) suggests a pre-Pleistocene origin of the major central European populations. *Molecular Ecology*, **9**, 709-722.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479-491.

Farris SF, Kallersjö M, Kluge AG, Bult C (1995) Constructing a significance test for incongruence. *Systematic Biology,* **44**, 570-572.

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, **17**, 368-376.

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

Fu Y, Li W (1993) Statistical tests of neutrality of mutations. *Genetics,* **133**, 693-709.

Giuffra E, Bernatchez L, Guyomard R (1994) Mitochondrial control region and protein coding genes sequence variation among phenotypic forms of brown trout *Salmo trutta* from northern Italy. *Molecular Ecology,* **3**, 161-171.

Guinand B, Durand J, Laroche J (2001) Identifying main evolutionary mechanisms shaping genetic variation of *Leuciscus cephalus* L. 1758 (Cyprinidae) in Western Greece: discordance between methods. *Life Sciences,* **324**, 1045-1060.

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

Harrison AC (1952) The cape kurper (with notes on the rockey of the eastern province). *Piscator*, **23**, 82-91.

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by the molecular clock of mitochondrial DNA. *Journal of Molecular Evolution,* **22**, 160-174.

Hendey QB (1983) Cenozoic geology and palaeogeography of the fynbos region. In: Fynbos palaeoecology: a preliminary synthesis (Deacon HJ, Hendey QB and Lambrechts JJN, eds). South African National Scientific Programmes report 75. Pretoria: CSIR.

Impson ND, Bills IR, Cambray JA (2002) A conservation plan for the unique and highly threatened freshwater fishes of the Cape Floral Kingdom. In: *Conservation of freshwater fishes: options for the future* (eds. Collares-Pereira MJ, Cowx IG, Coelho MM), pp. 432-440. Fishing News Books, London.

Kimura M, (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution,* **16**, 111-120.

Kocher TD, Thomas WK, Meyer A, Edwards SV, Paabo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the USA*, **86**, 6196-6200.

Liem KF (1963) The comparitive osteology and phylogeny of the Anabantoidei (Teleostei, Pisces). *Illinois Biological Monographs*, **30**, 1-149.

Machordom A, Doadrio I (2001) Evidence of a cenozoic betic–kabilian connection based on freshwater fish phylogeography (*Luciobarbus,* Cyprinidae). *Molecular Phylogenetics and Evolution*, **18**, 252-263.

Maddison WP, Maddison DR (1999) MacClade: Analysis of phylogeny and character evolution. Version 3.08. Sinauer Associates, Sunderland, Massachusetts.

Marjoram P, Donnely P (1994) Pairwise comparisons of mitochondrial DNA sequences in subdivided populations and implications for early human evolution. *Genetics,* **136**, 673-683.

Meyer A (1994) Shortcomings of the cytochrome b gene as a molecular marker. *Trends in Ecology and Evolution,* **9**, 278-280.

Moritz C (1994a) Applications of mitochondrial DNA analysis in conservation: critical review. *Molecular Ecology*, **3**, 401-411.

Moritz C (1994b) Defining evolutionarily significant units for conservation. *Trends in Ecology and Evolution*, **9**, 373-375.

Nedbal MA, Phillip DP (1994) Differentiation of mitochondrial DNA in largemouth bass. *Transactions of the American Fisheries Society*, **123**, 460-468.

Nei M, Li W (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the USA*, **76**, 5269-5273.

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

Norris SM, Teugels GG (1990) A new species of *Ctenopoma* (Teleostei: Anabantidae) from south eastern Nigeria. *Copeia*, **1990**, 472-479.

Norris SM (1994) The osteology and phylogenetics of the Anabantidae (Osteichthys, Perciformes). PhD dissertation, Arizona State University.

Orti G, Bell MA, Reimchen TE (1994) Global survey of mitochondrial DNA sequences in the threespine stickleback: evidence for recent migrations. *Evolution,* **48**: 608-622.

Palumbi SR, Martin AP, Romano S, McMillan WO, Stice L, Grabowski G (1991) "The simple fool's guide to PCR, version 2.0." Published and distributed by the authors, Honolulu, HI.

Park JY, Lee HJ, Kim WJ, Lee JH, Min KS (2000) Mitochondrial cytochrome *b* sequence variation in Korean salmonids. *Journal of Fish Biology,* **56**, 1145-1154.

Pennock DS, Dimmick WW (1997) Critique of the evolutionarily significant unit as a definition for 'distinct population segments' under the U. S. Endangered Species Act. *Conservation Biology*, **11**, 611-619.

Perdices A, Bermingham E, Montilla A, Doadrio I (2002) Evolutionary history of the genus *Rhamdia* (Teleostei: Pimelodidae) in Central America**.** *Molecular phylogenetics and Evolution*, **25**, 172-189.

Perdices A, Doadrio I, Economidis PS, Bohlen J, Banarescu P (2003) Pleistocene effects on the European freshwater fish fauna: double origin of the cobitid genus *Sabanejewia* in the Danube basin (Osteichthyes: Cobitidae). *Molecular Phylogenetics and Evolution*, **26***,* 289-299.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics Applications Note. **14**, 817-818.

Posada D, Crandall KA, Templeton AR (2000) GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487-488.

Ptacek MB, Breden F (1998) Phylogenetic relationships among the mollies (Poeciliidae: *Poecilia*: *Mollienesia* group) based on mitochondrial DNA sequences. *Journal of Fish Biology,* **53**, 64-81.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution,* **9**, 552-569.

Rogers AR (1995) Genetic evidence for a Pleistocene population expansion. *Evolution*, **49**, 608-615.

Ryder OA (1986) Species conservation and systematics: the dilemma of subspecies. *Trends in Ecology and Evolution*, **1**, 9-10.

Saiki RK, Gelfand DH, Stoffel SJ, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA, (1988) Primer directed enzymatic amplifications of DNA with a thermostable DNA polymerase. *Science,* **239,** 487-491.

Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution,* **4**, 406-425.

Sambrook J, Russell DW (2001) Molecular cloning a laboratory manual. Third Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

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

Seisser WG, Dingle RV (1981) Tertiary sea-level movements around Southern Africa. *Journal of Geology,* **89**, 83-96.

Simonsen KL, Churchill GA, Aquadro CF (1995) Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics,* **141**, 413-429.

Skelton PH (1980a) Systematics and biogeography of the redfin *Barbus* species (Pisces: Cyprinidae) from southern Africa. Phd Dissertation, Rhodes University.

Skelton PH (1980b) Aspects of freshwater fish biogeography in the eastern Cape. *The Eastern Cape Naturalist,* **24**, 17-22.

Skelton PH (1993) A complete guide to the freshwater fishes of South Africa. First Edition. Southern Book Publishers (Pty) Ltd, Halfway House, South Africa.

Swofford DL (1998) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

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

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution,* **10**, 512-526.

Tankard JA (1975) Cenozoic sea-level changes: a discussion. *Proceedings of the South African Society of Quaternary Research,* **71**, 1-17.

Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and analysis of alcohol dehydrogenase activity in Drosophila. *Genetics*, **117**, 343-351.

Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics*, **134**, 659-669.

Templeton AR, Routman E, Phillips CA (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.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research,* **22**, 4673-4680.

Van Wyk AE, Smith GF (2001) Regions of floristic endemism in Southern Africa: a review with emphasis on succulents. UMDAUS Press, Hatfield, South Africa.

Waples RS (1998) Evolutionarily significant units, distinct population segments, and the endangered species act: reply to Pennock and Dimmick. *Conservation Biology*, **12**, 718-721.

Appendix I

Locality information of all *Sandelia capensis* populations sampled during field surveys and the names of the collectors.

Population (River)	River system	Co-ordinates	Collectors**
Langvlei	Langvlei	32° 12' 30" S* 18° 25' 30" E	ES, IR
Verlorenvlei	Verlorenvlei	32° 30' 10" S* 18° 35' 50" E	DI
		32° 35' 50" S 18° 45' 00" E	ES, IR
		32° 28' 30" S 18° 32' 30" E	ES, IR
Boesmans	Berg	32° 46' 30" S* 18° 35' 00" E	DI
		32° 46' 30" S 18° 35' 15" E	ES, IR
Wemmershoek	Berg	33° 50' 28" S 19° 07' 03" E	ES, AC
Slanghoek	Berg	33° 36' 20" S 19° 12' 50" E	ES, HR
Diep	Diep	33° 31' 07" S* 18° 40' 30" E	ES, IR
Palmiet	Palmiet	34° 04' 30" S* 19° 02' 30" E	ES, HR
Nuwejaars	Nuwejaars	34° 31' 20" S 19° 42' 40" E	ES, HR
Grashoek	Heuningnes	34° 34' 15" S* 19° 56' 45" E	ES, HR
Grashoek	Heuningnes	34° 35' 20" S 19° 58' 30" E	ES, HR
Wolwekloof	Breede	33° 33' 50" S* 19° 07' 55" E	DI
Witte	Breede	33° 34' 30" S* 19° 08' 30" E	ES, HR
Dwars	Breede	33° 21' 30" S 19° 17' 00" E	ES, HR
		33° 21' 30" S 19° 18' 00" E	ES, HR
Groot	Breede	33° 35' 30" S 19° 19' 45" E	ES, HR
Cogmanskloof	Breede	33° 46' 20" S 20° 07' 10" E	ES, HR
		33° 47' 00" S 20° 07' 05" E	ES, HR

Keisers§	Breede	33° 59' 30" S 19° 47' 30" E	ES, HR
Poesjenels	Breede	33° 58' 50" S 19° 32' 40" E	ES, HR
Nuy	Breede	33° 37' 45" S 19° 41' 00" E	ES, HR
Genadendal	Breede	34° 02' 10" S 19° 33' 30" E	ES, HR
Greyton	Breede	34° 02' 20" S 19° 37' 10" E	ES, HR
Slang	Breede	34° 07' 30" S 19° 48' 30" E	ES, HR
Bok	Breede	34° 07' 10" S 19° 51' 10" E	ES, HR
Olifants	Breede	34° 07' 20" S 19° 55' 30" E	ES, HR
“East of Groenkloof”	Breede	34° 05' 50" S 20° 02' 20" E	ES, HR
“Groen-Jongenskloof”	Breede	34° 06' 30" S 20° 00' 30" E	ES, HR
Milkwood farm dam	Breede	34° 22' 40" S 20° 40' 40" E	ES, HR
Leeu	Breede	34° 00' 00" S 20° 20' 00" E	ES, HR
Heritage farm lower site	Breede	33° 59' 30" S 20° 25' 30" E	ES, HR
Melkhout	Breede	34° 22' 00" S 20° 38' 20" E	ES, HR
Duiwenhoks	Duiwenhoks	34° 05' 30" S 20° 57' 40" E	ES, HR
Kleinkruis	Goukou	34° 00' 52" S* 21° 17' 24" E	ES, HR, RvN
Langtou	Gourits	33° 58' 30" S* 21° 47' 20" E	ES, HR, RvN
		33° 58' 30" S 21° 46' 50" E	ES, HR, RvN
Bos	Gourits	33° 43' 50" S* 21° 30' 22" E	ES, HR, RvN
Hoekplaas	Gourits	33° 34' 00" S* 22° 58' 00" E	ES, HR, RvN
Groot Doring	Gourits	33° 47' 47" S 22° 15' 30" E	ES, HR, RvN
Weyers	Gourits	34° 01' 29" S 21° 35' 00" E	ES, HR, RvN
Groot	Gourits	33° 16' 20" S 22° 21' 15" E	ES, TB
		33° 17' 42" S 22° 20' 47" E	ES, TB

Nels	Gourits	33° 28' 03" S 21° 44' 09" E	ES, AC, TB
Kruis	Gourits	33° 28' 45" S 21° 54' 10" E	ES, AC, TB
Klein Brak	Klein Brak	33° 57' 10" S 21° 58' 45" E	ES, HR, RvN
Groot Brak	Groot Brak	33° 58' 38" S* 22° 11' 38" E	ES, HR, RvN
		33° 58' 24" S 22° 12' 19" E	ES, HR, RvN
		33° 57' 39" S 22° 14' 18" E	ES, HR, RvN
Gwaing	Gwaing	33° 57' 30" S 22° 18' 49" E	ES, HR, RvN
Moeras	Gwaing	33° 57' 30" S 22° 21' 27" E	ES, HR, RvN
Malgas	Malgas	33° 56' 14" S* 22° 25' 19" E	ES, HR, RvN
Touws	Touws	33° 58' 40" S* 22° 36' 20" E	ES, HR, RvN, RU
Duiwe	Duiwe	33° 59' 00" S 22° 39' 00" E	ES, HR, RvN, RU
Karatara	Karatara	33° 58' 55" S* 22° 50' 11" E	ES, HR, RvN, HI
		33° 59' 54" S 22° 49' 31" E	ES, AC
Piesang	Piesang	34° 03' 30" S 23° 22' 30" E	ES, HR, RvN
Bos	Bietou	34° 01' 29" S 23° 16' 12" E	ES, HR, RvN
		34° 00' 15" S 23° 20' 04" E	ES, HR, RvN
Farm dam (Stofpad Farm)	Bietou	33° 59' 00" S 23° 17' 00" E	ES, HR, RvN
Jakkalskraal	Bietou	34° 03' 59" S 23° 18' 54" E	ES, HR, RvN
Keurbooms	Keurbooms	33° 56' 17" S* 23° 22' 00" E	ES, HR, RvN
		33° 53' 22" S 23° 15' 46" E	ES, HR, RvN
Tsitsikamma	Tsitsikamma	34° 05' 49" S 24° 26' 52" E	ES, HR, RvN
Gheis	Kabeljous	33° 56' 26" S* 24° 48' 17" E	ES, HR, RvN
Witte	Gamtoos	33° 39' 13" S 24° 32' 06" E	ES, HR, RvN

Braam	Gamtoos	33° 43' 58" S* 23° 57' 40" E	ES, HR, RvN
Maitlands	Maitlands	33° 57' 30" S* 25° 16' 30" E	ES, HR, RvN
		33° 57' 30" S 25° 21' 00" E	ES, HR, RvN
Baakens	Baakens	33° 57' 32" S* 25° 30' 53" E	JAC
Blindekloof	Swartkops	33° 40' 20" S 25° 19' 00" E	ES, HR, DI, JA, TB, PB, RvN
Homtini	Homtini	33° 57' 30" S ¶ 22° 55' 05" E	ES, AC
		33° 57' 00" S 22° 55' 05" E	ES, AC
Gouna	Knysna	33° 59' 27" S 23° 02' 23" E	ES, AC
Hoogekraal	Hoogekraal	33° 58' 30" S 22° 48' 00" E	ES, AC
Wemmershoek	Berg	33° 50' 28" S 19° 07' 03" E	ES, GK

*Populations included in this study (see Table 1, Chapter 2).
**ES, Ernst Swartz; IR, Isa-Rita Russo; DI, Dean Impson; HR, Heidi Roos; RvN, Roelien van Niekerk; AC, Amanda Curtin; JAC, Jim A. Cambray; JA, John Allen; TB, Tom Barry; PB, Paulette Bloomer; RU, I.R. Russel; HI, H. Herd; GK, G. Kempen.
§*S. capensis* translocated from Hermanus.
¶*S. capensis* was only seen at this locality, not sampled.

Appendix II

Mitochondrial DNA control region haplotypes (369bp in length) of 21 populations of *Sandelia capensis*. The 31 unique haplotypes are named as in Table 1 of the materials and methods section (Chapter 2).

```
D8     AATGTATAAATGCATATATGCATATATGCATATATGTATTTACACCATAATTATAATTCTAGGGCATATGTAATGGATTACTAATATATATTAAACCTCAAAA-TATTAAGTATATGCCC-ATG
D9     C.G.............A.......A.....................................................AA.....CT.C.........G.........-C.....C.....A...-...
D10    C.G.............A.......A.....................................................AA......T.C.........G.........-C.....C.....A...-...
D11    C.AA.......A....C.....................................................A.................G..........-.......C....A...-...
D12    C...................T.....................................................A.....CC.....................-......A.....A.TT-...
D13    C.AA.......A....C.....................................................A.................G..........-............A...-G..
D14    C...................T.....................................................A.....CC.....................-......A.....A.TT-...
D15    C.....................................................................A.....CC.....................-.........AG.A.TA-...
D16    C.....................................................................A.....CC.....................A.........AG.A.TA-...
D17    C.....................................................................A.....CC.....................A.........AG.A.TA-...
D18    C...................T.....................................................-.....CC.........T..........-............A.TA-...
D19    C...................T.....................................................-A.....CC.C.................-............A.T.-...
D20    C...................T.....................................................-A.....CC.C.................-............A.T.-...
D21    C...................T.....................................................-A.....CC.C.................-............A.T.-..T
D22    C...................T.....................................................-A.....CC.C.................-............A.T.-...
D23    C...................T.....................................................-A.....CC.....................-............A.T.-...
D24    C...................T.....................................................-A.....CC.C.................-............A.T.-...
D25    C...................T.....................................................-A.....CC.C.................-............A.T.-...
D26    C...................T.....................................................AA.....CC.....................-C........A..A.A.-...
D27    C...................T.....................................................AA.....CC.........G..........-.........A..A.A.-...
D28    C.......-...........T.....................................................AA.....CC...C...............-............A.A.-...
D29    C.....................................................................AA.....CC.....................-C.......G...A.TA-...
D30    C.C.....G.......A.......A.....................................................A......CC.....................-........G...A.TA-...
D31    C.......G.......A.......A.....................................................A......CC.....................-........G...A.TA-...
D1     ----------------------A..A.....................C.G...........T...GT.....T......C...A..............T.-C.....A..A..A.TAA...
D2     ----------------------A..A.....................C.G...........T...GT.....T......C...A..............T.-C.....A..A..A.TAG...
D3     ----------------------A..A.....................C.G...........T...GT.....T......C...A..............T.-C.....A..A..A.TAA...
D4     ----------------------A..A.....................C.G...........T...GT.....T......C...A..............T.-C.....A..A..A.TAG...
D5     ----------------------A..A.....................C.G...........T...GT.....T......C...A..............T.-C.....A.....A.TAA...
D6     ----------------------A..A.....................C.G...........T...GT....AT......C...A..............T.-C.....A.....A.TAA...
```

```
D8      AAATAAATGACATAATCCTATTGATATTACATAAACCATAATATGTAAAATCAACACTAACTTATTTAAATGTTAAAGGATAGTCAAGTTTTAACAAAAAATTCATTAGTAAAGATATACCAGG
D9      ..............C.....................G.................TC.....................................................G.............
D10     ..............C.....................G.................TC.....................................................G.............
D11     ......G..............................G.....................................C.....................................................
D12     ......-.A.....................................................T.............................T.............................
D13     ....................................G.....................................C.................C.....................
D14     ......-.A.....................................................T.............................T.............................
D15     ........A.............T.................GG..................T.....................................G..........G.C..............
D16     ........A............T.................GG..................T.....................................G............C..............
D17     ........A............T.................GG..................T.....................................G............C..............
D18     ........A..............................G..................T.................................C..............................
D19     ........A.....................................................T...T..........A.....................T..............................
D20     ........A.....................................................T...T.................................C..............................
D21     ........A.....................................................T.T...T.................................C..............................
D22     ........A.....................................................T...T.................................T..............................
D23     ........A.....................................................T...T.................................T..............................
D24     ........A....G................................................T...T.................................T..............................
D25     ........A.....................................................T...T.................................C..............................
D26     ....................................GG..............T.....................................T..............................
D27     ........A.............................GG..............T...T.....................................................................
D28     ........A.............................GG..............T...T..........A..........A.....................................................
D29     ....G...A.............................G...........T.........T..............G.....................................G.............
D30     ....G...A.............................GG..........T....T...T...........C.....................................G.............
D31     ....G...A.............................GG..........T....T...T...........C.....................................G.............
D1      ..-..G.........AT.-.C.A...............GG..............T...T..............GC.A.-.C.T...C.....T..............C...........A.
D2      .....G.........AT.-.C.A...............GG..............T...T..............GC.A.-.C.T...C.....T..............C...........A.
D3      ..-..G.........AT.-.C.A...............GG..............T...T..............GC.A.-.C.T...C.....T..............C...........A.
D4      ..-..G.........AT.-.C.A...............GG..............T...T..............GC.A.-.C.T...C.....T..............C...........A.
D5      ..-............AT.....A...............G...............T...T..............C.A.-.C.T...C.....T..............C...........A.
D6      ..-............AT.....A...............GG..............T...T..............C.A.-.C.T...C.....T..............T...........A.
D7      ..-............GT.....A...............GG..............T...T..........A...GC.A.-.C.T.........T..............T...........A.
```

```
D8     ACTCAACATTATATTAATTTAACCATTAAGCCCAATAAGAACCGACCTAAAGTGATTTTTAATTGCATACTCTTATTGATGGTGAGGGACAATTATTTGTGGGGGTAGTACTCAGTGAATT
D9     ..........................................................................................................CT........
D10    ..........................................................................................................CT........
D11    ......................A...................................................................................CT......C.
D12    ..........................................................................................................CT........
D13    ......................G.....................................................G.............A..........T......C.
D14    ..........................................................................................G???????????????
D15    ..........................................................................................................CT........
D16    ..........................................................................................................CT........
D17    ..................................................................G.......................................CT........
D18    ......T.C...................................................................................................CT........
D19    ......T.C.G.................................................................................................CT........
D20    ......T.C.G.................................................................................................CT........
D21    ......T.C.G.................................................................................................CT........
D22    ......T.C.G.................................................................................................CT........
D23    ......T.C.G.................................................................................................CT........
D24    ......T.C.G.................................................................................................CT........
D25    ......T.C...................................................................................................???????????
D26    ......T........G.............................................................................................CT......??
D27    ......T........G.............................................................................................CT......C.
D28    ......T........G............................................................................................CT........
D29    ..............................................C.....................................................CC....A.........CT........
D30    ..............................................C.....................................................??????????
D31    ..............................................C.....................................................CT........
D1     .T....T........G...C.....................................................................................C.....T........
D2     .T....T........G...G.....................................................................................C..............
D3     .T....T........G...A.....................................................................................C..............
D4     .T............G...G.....................................................................................C..............
D5     .T....T........G...G.....................................................................................A.........T........
D6     .T....T........G...G.....................................................................................A..................
D7     .T....T...G....G.....................................................................................................
```

Appendix III

Percentage sequence divergence values (HKY85 distances) within and between *Sandelia capensis* phylogenetic groups.

Phylogenetic groups	Control region sequence divergence (%)	Cytochrome b sequence divergence (%)
Within *S. capensis*	0.27 – 15.38	0.16 – 9.42
Within south coast	0.27 – 9.22	0.16 – 3.1
Within west coast	0.29 – 4.94	0.22 - 2.24
South coast vs. west coast	9.85 – 15.38	6.69 – 9.42
Diep vs. west coastal lineages	3.97 – 4.94	2.06 – 2.24
Berg vs. Verlorenvlei/Langvlei	2.39 – 3.03	0.82 – 1.33
Verlorenvlei vs. Langvlei	0.59 – 1.19	0 – 0.67
Heuningnes vs. south coastal lineages	5.04 – 7.31	1.1 – 2.73
Between divergent Goukou lineages	6.7	1.43
Within Maitlands	5.81	2.72
Eastern vs. western clade within the southern group	0.29 – 8.25	1.12 – 2.91

Appendix IV

Mitochondrial DNA cytochrome b gene haplotypes (614bp in length) of 20 populations of *Sandelia capensis*. The 22 unique haplotypes are named as in Table 1 of the materials and methods section (Chapter 2).

```
C6      GACTATTCCTTGCAATACACTATACTTCAGACATCACAACAGCCTTTTCATCCGTTGCTC
C7      ............................................................
C8      ????????????????????????????????????????????????????????????
C9      ............................................................
C10     ...........................................................C....
C11     ????????????????????????????????????????????????????????????
C12     ............................................................
C13     ............................................................
C14     ............................................................
C15     ............................................................
C16     ?????????????????????????.....................................
C17     ..........A.................................................
C18     ............................................................C.
C19     ............................................................
C20     ............................................................
C21     ............................................................
C22     ............................................................
C1      ......................C..C..................................
C2      ......................C..C..................................
C3      ????????????????????????????????????????????????????????????
C4      ......................C..C..................................
C5      ?????????????????????.C.....T.....................................
```

```
C6      ATATCTGCCGGGATGTCAACCACGGCTGACTTATCCGAAATCTCCACGCCAATGGCGCTT
C7      ..............T.........................................C.
C8      ???????.......T.........................................
C9      ..............T.........................................
C10     ............C..T.........................................C.
C11     ????????????????????.....................................C.
C12     ..........A..C..T.........................................C.
C13     ............C..T.........................................C.
C14     ............C..T.........................................C.
C15     ............C..T.........................................C.
C16     ............C..T.........................................C.
C17     ............C..T.........................................C.
C18     ............C..T.........................................C.
C19     ............C..T.........................................C.
C20     ............C..T..T..A..............C..................C.
C21     ............C..T.....A..............C..................C.
C22     ............C..T.....A..............C..................C.
C1      ..........A.....T.....................C..................C.
C2      ..........A.....T.....................C..................C.
C3      ????????..A.....T.....................C..................C.
C4      ..........A.....T.....................C..................C.
C5      ..........A...........................C..................C.
```

```
C6      CTTTCTTCTTCATCTGCATTTATCTTCACATCGGACGGGGCCTCTATTATGGATCATACC
C7      ............................................................
C8      ............................................................
C9      ..................................G.........................
C10     ...........................................................G......
C11     ............................................................
C12     ............................................................
C13     ............................................................
C14     ............................................................
C15     ...........................................................G......
C16     ............................................................
C17     ...........................................................G......
C18     ............................................................
C19     ............................................................
C20     ..........................................................C.........
C21     ............................................................
C22     ............................................................
C1      ..........T.....................G..A..T...........C..C....
C2      ..........T.....................G..A..T...........C..C....
C3      ..........T.....................G..A..T...........C..C....
C4      ..........T.....................G..A..T...........C..C....
C5      ..........T........C.................A..T...........C..C....
```

```
C6      TTTATAAAGAGACATGAAACACAGGCGTAGTCCTTCTCCTTCTAGTTATAATAACTGCCT
C7      ............................A.T.........................
C8      ............................T.........................
C9      ............................T.........................
C10     ............................T.....T.....................
C11     ......................G.....T.....T.....................
C12     ................G.....G.....T.....T.....................
C13     ............................T.....T.....................
C14     ............................T.....T.....................
C15     ............................T.....T.....................
C16     ...........................A.T.....T.....................
C17     ............................T.....T......A..............
C18     ............................T.....T.....................
C19     ............................T.....T..C.....C............
C20     ............................T.....T.....................
C21     ............................T.....T.....................
C22     ............................T.....T.....................
C1      .C........A.................GA.T.....T..C...............
C2      .C........A.................GA.T.....T..C...............
C3      .C........A.................GA.T.....T..C...............
C4      .C........A.................G..T.....T..C...............
C5      .C........A.................GA.......T..C...............
```

```
C6      TTGTAGGCTATGTACTACCATGAGGACAAATATCATTTTGAGGCGCTACCGTCATTACAA
C7      ....G.......................................................
C8      ...........................................................
C9      ...........................................................
C10     ..........................................C................
C11     ....G.....................................C................
C12     ....G.....................................C................
C13     ..............T...........................C................
C14     ..........................................C................
C15     ..........................................C................
C16     ..........................................C................
C17     ..........................................C................
C18     ....G..A..................................C................
C19     ....G..A..................................C................
C20     ....G.....................................C................
C21     ....G.....................................C................
C22     ....G.....................................C................
C1      ....G..G........C.....................C........C..A.........
C2      ....G..G........C.....................C............A.........
C3      ....G..G........C.....................C............A.........
C4      ....G..G........C.....................C.....................
C5      ....G..G........C.....................C.....................
```

```
C6      ACCTCCTCTCCGCTGTCCCTTACATGGGGGACATGCTTGTCCAATGGATCTGAGGGGGCT
C7      ............................................................A....
C8      ...........................................................
C9      ...........................................................
C10     ...........................................................
C11     ............................A..............................
C12     ............................A..............................
C13     ...........................................................
C14     ...........................................................
C15     ...........................................................
C16     ...........................................................
C17     ............................A..............................
C18     ..........................................A................
C19     ........................A.....................A.............
C20     ...........................................................
C21     ...........................................................
C22     ...........................................................
C1      ..........T........C.....A..A........C.....................
C2      .......A..T........C.....A..A........C.....................
C3      .......A..T........C.........A........C.....................
C4      ..........T........C.....A..A........C.....................
C5      ..........T........C.....A..A........C................A....
```

```
C6      TTTCAGTAGACAACGCCACCCTCACCCGATTCTTCACCTTTCACTTCCTACTGCCCTTTA
C7      ............................................................
C8      ............................................................
C9      ............................................................
C10     ............................................................
C11     ...........................................................G........
C12     ...........................................................G........
C13     ............................................................
C14     ............................................................
C15     ............................................................
C16     ............................................................
C17     ............................................................
C18     ............................................................
C19     ............................................................
C20     ............................................................
C21     ............................................................
C22     ............................................................
C1      .C.........................G.....................A......G
C2      .C.........................G.....................A......G
C3      .C.........................G.....................A......G
C4      .C.........................G.....................A.......
C5      .C..........................................................A......G
```

```
C6      TTATCGCAGCCATAGTTATTATTCACCTACTATTTCTACATGAAACAGGCTCAAACAATC
C7      ............G...............................................
C8      ............G...............................................
C9      ............G.....................G.....................
C10     ............G...............................................
C11     ............G...............................................
C12     ............G...............................................
C13     ............G...............................................
C14     ............G...............................................
C15     ............G...............................................
C16     ............G...............................................
C17     ............................................................
C18     ............G...............................................
C19     ............G...............................................
C20     .C..........G...............................................
C21     .C..........G...............................................
C22     .C..........G...............................................
C1      ........................T..G..T..C..G...................C.
C2      ........................T..G..T..C..G...................C.
C3      ........................T..G..T..C..G...................C.
C4      .........................G..T..C..G...................C.
C5      .........................G..T..C..G...................C.
```

```
C6      CAATCGGACTAAACCCAAATGTAGACAAAATTTCATTCCACCCCTACTTCTCGTACAAAG
C7      ............................................................
C8      .G..........................................................
C9      ............................................................
C10     ............................................................
C11     ............................................................
C12     ............................................................
C13     ............................................................
C14     ....T.......................................................
C15     ....T.......................................................
C16     ....T.......................................................
C17     ............................................................
C18     ....T.......................................................
C19     ....T.......................................................
C20     ...................................T........................
C21     ..G................................T........................
C22     ...................................T........................
C1      ..............................C.............................
C2      ..............................C.............................
C3      ..............................C.............??????????????
C4      ..............................C.............................
C5      ..............................C.............................
```

```
C6      ATATTCTGGGGTTTGCAGCACTTCTAATTGCACTAACTACGCTCTCACTATTCTCCCCAA
C7      ............C.........A.....................................
C8      ............C.............................................
C9      ............C...................??????????????????????????
C10     ............C.............................................
C11     ..........A..C............G.................................
C12     ..........A..C............G.................................
C13     ...C......C..C.............................................
C14     ............C.............................................
C15     ............C.............................................
C16     ............C.............................................G.
C17     ..........A..C.............................................
C18     ..........T..C........................C....................
C19     ..........T...........................C..............T.......
C20     ..........T..C.............................A..................
C21     ..........T..C.............................................?
C22     ..........T..C.............................................
C1      .C..C........C..................G........A.....C............
C2      .C..C........C..................G........A.....C............
C3      ????????????????????????????????????????????????????????
C4      .C..C........C..................G........A.....C............
C5      .C..C..A.....C.........................A..................
```

```
C6      ACTTACTGGGAGAC
C7      ..............
C8      ..............
C9      ??????????????
C10     ...???????????
C11     .......A......
C12     .......A......
C13     ..............
C14     ..............
C15     ..............
C16     ..............
C17     ..............
C18     ..............
C19     .???????????????
C20     ..............
C21     ??????????????
C22     ..............
C1      .......A......
C2      .......A......
C3      ??????????????
C4      .......A......
C5      .......A......
```

Appendix VI

Mitochondrial 16S ribosomal RNA gene haplotypes (486 bp) of five species (four genera) of the family Anabantidae. Haplotype designations follow Table 1. A Gourami was used as an outgroup.

```
A1          AGTGACTTAGTTCAACGGCCGCGGTATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTC
A2          ........G.....................................................
A3          ......AAC...T.................................................
A4          ......AAC...T.................................................
A5          T.....AA....T..............C...................................
A6          ......AA....T.................................................
A7          ?.....AA....T.................................................
A8          ????????????????.............................................
A9          ?.....AA....T.................................................
A10         ......AA....T.................................................
GOURAMI     .....TGA....T.....................................T......C....

A1          TTTTAAATGAAGACCTGTATGAATGGCAAGACGAGGGCTTAACTGTCTCCTTTTTCAAGT
A2          ............................................................
A3          ........................C....C............................CG..
A4          ........................C....C............................CG..
A5          ..............................TA............................TG..
A6          ..............................C..........A..................A.
A7          ..............................C.............................
A8          ..............................C............G.................
A9          ..............................C.............................
A10         ..............................C............G.................
GOURAMI     ..............................TA..........T..................

A1          CAATGAAATTGATCTCCCCGTGCAGAAGCGGGGATACTACCATAAGACGAGAAGACCCTA
A2          ....................................T.......................
A3          ................T..............A....AA.A.....................
A4          ................T..............A....AA.A.....................
A5          ...C............T..............A....AA.......................
A6          ................T...............A...AAC.......................
A7          ....A.........C.....................AAG.....................
A8          ....A.........C................A...AA.......................
A9          ....A.........C.....................AA.....................
A10         ....A.........C.....................AA.....................
GOURAMI     ................T............A....A....C.....................

A1          TGGAGCTTTAGACACTAAAGCAGCTCTTGTCAATACCCTAAGAATAATTAGACTAAACTT
A2          ............................................................
A3          ..............G.......A......T..A....CT.A..C..A-...T.....CA
A4          ..............G.......A......T..A....CT.A..C..A-...T.....CA
A5          ....................T..AC.........CC....T.A.....A-.A.TCG...AA
A6          ..............G...A...A....AATGTAC..GCT.A..C..A-....CG..TCA
A7          ..............G.......T......T..GC.T.CT.A..C..A-..G.C....CA
A8          ..............G.......A......T..AC...CT.A..C..A-....CT...CA
A9          ..............G.......T......T..GC.T.CT.A..C..A-..G.C....CA
A10         ..............G.......A......T..AC...CT.A..C..A-....C....CA
GOURAMI     ......................A......T..ATT..CTTA.....A-...T.....CA
```

```
A1          TGGG-AACCCTGCTTCAATGTCTTTGGTTGGGGCGACCACGGGGAAGTAAGTAACCCCCG
A2          ....-.......................................................
A3          AAAAT..A......AT.....................G.......AC..AC.......A
A4          AAAAT..A......AT.....................G.......AC..AC.......A
A5          AAA.-GC....A.CCT.....................G.......A...AA.......A
A6          GAA.-..G....TCCT.....................G.......AC..AA.T.....A
A7          AAA.-..A......AT.....................G.......AC..AA.......A
A8          AAA.-..G......AT.....................G.......AC..AA.......A
A9          AAA.-..A......GT.....................G.......AC..AA.......A
A10         AAA.-..G......AT.....................G.......AC..AA.......A
GOURAMI     AAA.-..A......CT..C..................G.......AC..A........A
```

```
A1          CGCGGAATAAGAGAACATCTCTCGCAACCAAGAGCTTCTGCTCTAAGTAACAGAATATCT
A2          ............................................................
A3          ..T....CG....T.TT......A..G........CA.A.......T.........T...
A4          ..T....CG....T.TT......A............CA.A.......T.........T...
A5          T.T...GA.........C.....TA...A.C.....CA.A.......T.............
A6          ..T....CG...CC..C.....TA..........TAA.A.......T.........C....
A7          A.T.....G...AT..CC.....A............CA.A.......T.........C....
A8          A.T.........AT..C.T....A............CA.A.......T.........C....
A9          A.T.....G...AT..CC.....A............CA.A.......T.........C....
A10         A.T.........AT..C.T....A............CA.A.......T.........C....
GOURAMI     T.T...CCG...AC..TGT...TAA..........CA.A.......T.........T...
```

```
A1          GACCATAACGATCCGGTAAGACCGATCAACGGACCGAGTTACCCTAGGGATAACAGCGCA
A2          ............................................................
A3          ........-.......C..CG.......................................
A4          ........-.......C..CG.......................................
A5          ....T...-.......C..TG.......................................
A6          ........-.......C..TG.......................................
A7          .......T-.......C..CG.......................................
A8          .......T-.......C.GCG.......................................
A9          .......T-.......C..CG.......................................
A10         .......T-.......C.GCG.......................................
GOURAMI     ....T...-.......C..CG.......................................
```

```
A1          ATCCTCTTTTAGAGTTCGTATCGACAAGAGGGTTTACGACCTCGATGTTGGATCAGGACA
A2          ............................................................
A3          ....C.........CC.A..........G................................
A4          ....C.........CC.A..........G................................
A5          ....C.........CC.A..........G................................
A6          ....C.........CC.A..........G...C............................
A7          ....C.........CC.A..........G................................
A8          ....C.........CC.A..........G................................
A9          ....C.........CC.A..........G................................
A10         ....C.........CC.A..........G................................
GOURAMI     ..............CC.A..........G................................
```

```
A1          TCCTAA
A2          ......
A3          ......
A4          ......
A5          ......
A6          ......
A7          ......
A8          ??????
A9          ......
A10         ......
GOURAMI     .....T
```

Appendix VII

Mitochondrial DNA control region haplotypes (339 bp) of seven populations of *Sandelia bainsii*. Haplotype designations follow Table 2.

```
B1      TTCTCTGAATATTTTACATATATGTATTTACACCATCAAATTTAGAGCATTTCTATTCAA
B2      ............................................................
B3      ............................................................
B4      ...........................................G................

B1      TATATACATTAATATATTAATACCCATTACCGAGATAGAA-CATCAGATTTTTTAGTATA
B2      ...........................................-................
B3      ...........................................-................
B4      ........................C..T.C......A..CTG.....A.........

B1      AGTTACTAGGGGATAACATAAGATCTAATATGTAGAAACAATATTAAATGTAAACACACC
B2      ............................................................
B3      ............................................................
B4      .A.C.........................A..........................

B1      TGGCGAAATTTAAGATCTATCAAAAAACTCATTAGTCTAGATATACCAAGAAACACCATG
B2      ............................................................
B3      ............................................................
B4      ............................................................A

B1      GTATTAATTAAACCAATAAGCCCAATAAGAACCGACCTAAAGTGATTTTTAGGCGCATAC
B2      ............................................................
B3      ............................................................
B4      A.....G........G..T..............................A........

B1      GGTTATTGAAGGTGAGGGACAATTATTTGTGGGGGTCAC
B2      ............................G.........
B3      ....................G.................
B4      ............................A...........
```