Development of microsatellite markers in the Sparidae and their application in population genetics of the hottentot seabream *Pachymetopon blochii* (Valenciennes, 1830) along the coast of South Africa

By

Kerry Reid

Submitted in partial fulfilment of the requirements for the degree

MAGISTER SCIENTIAE

in the Faculty of Natural and Agricultural Sciences
University of Pretoria

Pretoria
March 2011

Development of microsatellite markers in the Sparidae and their application in population genetics of the hottentot seabream *Pachymetopon blochii* (Valenciennes, 1830) along the coast of South Africa

By

Kerry Reid

Supervisor: Prof. P. Bloomer
Molecular Ecology and Evolution Programme
Department of Genetics
University of Pretoria
Pretoria

Co-supervisor: Dr. T. B. Hoareau
African Coelacanth Ecosystem Programme
Department of Genetics
University of Pretoria
Pretoria

Department of Genetics

Magister Scientiae

Dissertation abstract

The marine environment is vast and is one of the last natural habitats that provides' a major source of protein through fisheries. This has led to a massive overexploitation of this resource because in the past marine fish populations were perceived as inexhaustible. Recently it has been observed that this is not the case, with many fish populations collapsing which has initiated further investigation of these populations. By combining different aspects of research (species biology, movement and genetics) more complete management and conservation plans may be implemented. Marine studies have greatly benefited from the advancement of molecular tools that allow for

inferences to be made about gene flow and connectivity between habitats which could not previously be made as it is not easy to monitor marine species within their environment.

South Africa has a wide range of marine fish species due to the diverse geological and oceanographic features which provide a number of habitats along the coast. One of the most prevalent and diverse fish families along the South African coast is the Sparidae. This family is of economic importance to the line fishery and many of these species are considered vulnerable or endangered due to a combination of overfishing and life-history traits. Two species of particular interest are the hottentot seabream (*Pachymetopon blochii*) and white steenbras (*Lithognathus lithognathus*). Both species are endemic to southern Africa and are considered vulnerable to overexploitation.

This study reports the development of microsatellite markers for both the above mentioned species using the FIASCO protocol and next generation 454 sequencing. Nine polymorphic markers were identified in each of these species. Nineteen markers (fifteen newly developed and four from other sparids) were used as a panel on eleven economically important sparids, to test microsatellite cross-species amplification. From this study, I was able to identify twelve adequate polymorphic loci for the white steenbras (applied in a population genetic study by a PhD student associated with SAIAB) and fourteen in the hottentot seabream which were applied in this dissertation. We were also able to identify a number of polymorphic loci for the other sparids. It was concluded that the sparids do not show a negative correlation between genetic distance and microsatellite amplification success and polymorphism.

The study further investigated the population genetic structure of the hottentot seabream. This species is thought to have sedentary adults and a relatively patchy habitat distribution. Using the above mentioned markers on individuals sampled in 2001 and 2009, we were able to identify a weak spatial and temporal variation. It was concluded that the observations of weak population structure were likely due to variation in reproductive success and environmental patchiness which led to changes observed between sampling years.

Declaration

I, ... declare that this dissertation, which I hereby submit for the degree ..... at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signature: ……. Date: …….

Acknowledgements

Firstly I would like to thank my supervisors Prof. Paulette Bloomer and Dr Thierry Hoareau for their guidance. Without their knowledge and encouragement I would have been lost. Thank you, Prof. Bloomer for your invaluable advice, support and funding for this project. Dr Hoareau I would like to thank you for all the constructive criticism and many discussions that have definitely improved me as a scientist.

I would like to thank Marine and Coastal Management (MCM) for the hottentot seabream (*Pachymetopon blochii*) samples collected in 2001 and Steve Benjamin who collected samples for this species in 2009. I would also like to thank Rhett Bennett for the white steenbras (*Lithognathus lithognathus*) samples and the South African Institute for Aquatic Biodiversity (SAIAB) for the other sparid samples required for this study. I would also like to thank the African Coelacanth Ecosystem Programme (ACEP) for funding and the University of Pretoria for providing a wonderful working environment.

I would especially like to thank my parents (Ann and William) and family who have always encouraged me to study and follow my passion. I would also like to thank my friends (especially those in the Department of Genetics) who always lent me a shoulder to cry on when things were not working out and encouragement when they were.

Finally, I would like to thank my Molecular Ecology and Evolution Programme (MEEP) lab members for helping to form a pleasant learning and working environment. I always enjoyed interacting with everyone, sharing ideas and having constructive discussions.

Table of contents

Title page..... I
Thesis abstract..... II
Declaration..... IV
Acknowledgements..... V
Table of Contents..... VI
List of Tables..... X
List of Figures XI
List of Appendices..... XIII

Chapter 1: 1

General Introduction..... 1

Abstract..... 1
1.1 Introduction..... 2
1.2 Marine management and conservation in the marine realm..... 3
1.3 Life-history traits and marine factors influencing population genetics in marine fish..... 4
1.4 Molecular Markers..... 8
1.4.1 Microsatellites..... 9
1.4.2 Cross-species amplification..... 11
1.5 South African marine context..... 12
1.5.1 Line fisheries around South Africa..... 13
1.6 Coastal fishes of southern Africa..... 14
1.6.1 The fish fauna..... 15
1.6.2 The family Sparidae..... 15
1.6.3 *Lithognathus lithognathus*..... 16
1.6.4 *Pachymetopon blochii*..... 16
1.7 Aim and Objectives..... 18
1.8 Dissertation outline..... 20
1.9 References..... 22

Chapter 2: 30

Isolation and characterization of microsatellites in two South African endemic sparids *Pachymetopon blochii* (Valenciennes 1830) and *Lithognathus lithognathus* (Curvier 1829) and cross-species amplification in nine other economically important sparids

Abstract 30
2.1 Introduction 31
2.2 Materials and Methods 33
2.2.1 Sampling and DNA extraction 33
2.2.2 Development of microsatellite markers in the hottentot seabream and white steenbras 33
2.2.3 Sequencing of homozygotes 35
2.2.4 Cross-species amplification 35
2.2.5 Data analysis of microsatellites 37
2.2.6 Evaluation of cross-species amplification and polymorphism 38
2.3 Results 38
2.3.1 Sequencing results from 454 sequencing 38
2.3.2 Isolation and characterization of microsatellites 39
2.3.3 Sequencing of homozygotes 40
2.3.4 Cross-species amplification 43
2.3.5 Correlation between genetic distance and amplification success and polymorphism 43
2.4 Discussion 47
2.4.1 Development of microsatellite with 454 pyrosequencing 47
2.4.2 Cross-species amplification and correlation testing 49
2.4.3 Conclusion 51
2.5 References 52
2.6 Appendices 56

Chapter 3: 64

Genetic differentiation in the hottentot seabream *Pachymetopon blochii* (Valenciennes, 1830) from South Africa based on microsatellite variation

Abstract..... 64
3.1 Introduction..... 65
3.2 Materials and methods..... 68
3.2.1 Sampling..... 68
3.2.2 Molecular markers..... 69
3.2.3 DNA extraction and microsatellite amplification..... 70
3.2.4 Summary statistics..... 71
3.2.5 Inferring the number of populations..... 71
3.2.6 Analysis of molecular variance (AMOVA)..... 72
3.2.7 Factorial correspondence analysis (FCA)..... 72
3.2.8 Isolation by distance (IBD)..... 73
3.2.9 Spatial autocorrelation (SAC)..... 73
3.2.10 Effective population size (N_e)..... 74
3.3 Results..... 74
3.3.1 Summary statistics..... 74
3.3.2 Inference of the number of populations..... 77
3.3.3 Analysis of molecular variance..... 78
3.3.4 Factorial correspondence analysis..... 78
3.3.5 Isolation by distance..... 81
3.3.6 Spatial autocorrelation..... 81
3.3.7 Effective population size..... 83
3.4 Discussion..... 84
3.4.1 Population identification..... 84
3.4.2 Spatial genetic variation..... 85
3.4.3 Temporal genetic variation..... 87
3.4.4 Conclusion..... 87
3.5 References..... 88
3.6 Appendices..... 92

Chapter 4: 97

Dissertation synthesis

4.1 Future work..... 98

4.2 References..... 98

List of Tables

2.1A Isolated microsatellite loci from 50 *Pachymetopon blochii* samples with characterization of the primer sequence, repeat motif, allele size range, Observed (H_O) and Expected (H_E) heterozygosity and F_{IS} (Weir & Cockerham, 1984)..... 41

2.1B Isolation of microsatellites from 50 white steenbras (*Lithognathus lithognathus*) samples, including primer sequence, repeat motif, allele size range, number of alleles (N_A), Observed (H_O), Expected hetrozygosity (H_E) and F_{IS} (Weir and Cockerham, 1984)..... 42

2.2 Number of alleles (Na), Observed (Ho) and Expected (He) heterozygosity and F_{IS} (Weir & Cockerham, 1984) in eleven sparid species found around South Africa..... 44

3.1 A summary of the sampling locations, geographic co-ordinates, year sampled and number of individuals sampled..... 70

3.2 “Standard errors of the von Bertalanffy growth parameters and the lengths at age, calculated by the jack-knife method of variance estimation” taken from Pulfrich and Griffiths (1988b)..... 73

3.3 Pairwise F_{ST} (Raufaste and Bonhomme 2000) and R_{ST} values between locations and years; F_{ST} values are above and R_{ST} values are below the diagonal. Blue shading indicates the pairwise comparisons between 2001 locations, purple indicates pairwise comparisons between 2009 locations and pink cells are comparisons between years. Significant pairwise values are indicated in bold..... 75

3.4 The total number of alleles per location, A_R and the average H_O, H_E and F_{IS} over all markers. Deviations from HWE are indicated in bold..... 76

3.5 Summary of the H_O, H_E, A_R and F_{IS} over all markers for each size class in each year. Deviations from HWE are indicated in bold..... 77

3.6 A summary of Analysis of Molecular Variance (AMOVA). The table indicates the source of genetic variation (% GV), the degrees of freedom (d.f.) and the probability that the percent of molecular variance significantly deviates are indicated in bold..... 78

List of Figures

1.1 Population models differing by their level of population divergence from completely isolated sub-populations (A) to panmixia (D). A) Three isolated sub-populations, B) three sub-populations with low connectivity C) sub-populations with high connectivity and D) panmixia. Diagram from Waples & Gaggiotti (2006)..... 5

1.2 Map of the world (google maps) with a focus on South Africa. The enlarged map shows the biogeographical provinces and currents around South Africa. Spawning area (yellow) upwelling zones (green) are shown on the west coast. The "jet" current (black arrows) moves along the west coast and feeds into an inshore current (red arrow). Modified from Whitfield *et. al*. (2001)..... 14

1.3 Distribution range of the hottentot seabream (red) and the white steenbras (yellow). The dashed lines indicate where each species is thought to occur infrequently; the solid lines indicate the main distribution of the species..... 17

2.1 Sparid phylogeny based on the mitochondrial DNA cytochrome b gene. The coloured dots indicate the genera represented in the present cross-species amplification tests, their position in the phylogeny and the clade they belong to. The two focal genera are indicated by black arrows This figure was modified from Chiba *et. al.* (2009). The two main species included in this study were not included in the phylogeny..... 36

2.2 **A**) Distributions of size lengths of the fragments obtained from 454 LifeSciences/Roche GS-FLX sequencing for *P. blochii* (blue) and *L. lithognathus* (red) after repeat enrichment. **B**) Comparison of the number of sequences for each species containing six repeats or more of each repeat type..... 39

2.3 Relationship between the percentage sequence divergence between source species and target species and the percentage of amplification (AL) and polymorphism (PL) of the microsatellite loci. The trend lines represent the predicted negative correlation from the correlation equations of Carreras-Carbonell *et. al.* (2008) for percentage AL (blue) and PL (red) and the observed amplification success for **A**) *P. blochii* and **B**) *L. lithognathus* loci and polymorphism success **C**) *P. blochii* and **D**) *L. lithognathus* are shown (see Appendix 2E)..... 46

3.1 Map of the locations sampled along the coast of South Africa in 2001 and

2009. The 2001 sampling locations are indicated in blue and the 2009 sampling locations in red. The inset shows a map of Africa (google maps; orange line indicates the west coast of southern Africa where the hottentot is distributed) and an image of the hottentot seabream..... 69

3.2 Structure graphs for 2001 (**A**) and 2009 (**B**). In both graphs the left Y axis L (K) and the solid lines indicate the likelihood value of a specific number of populations (K) determined by STRUCTURE and the right Y axis and dashed trend lines indicate the *ad hoc* statistic Δ K..... 77

3.3 Factorial Correspondence Analysis between locations. **A**) 2001 locations and **B**) 2009 locations. Colours for locations on the map correspond to the colours in the FCA..... 79

3.4 Factorial Correspondence Analysis. **A**) 2001 (yellow) and 2009 (Blue) combined samples, **B**) 2001 samples and **C**) 2009 samples. In graphs B and C, yellow corresponds to 4-5 years, blue to 5-6 years, white to 6-7 years, grey 7-8 years, pink 8-9 years, turquoise 9-10 years and dark blue 10-11 years..... 80

3.5 Isolation by distance using linearized $F_{ST}/1-F_{ST}$ and geographic distance. The correlation coefficients were 0.064 (*P*-value > 0.05) and 0.020 (*P*-value > 0.05) for 2001 and 2009 respectively. The blue points indicate the data for 2001 and the red for 2009..... 81

3.6 Spatial autocorrelation graphs across size classes for 2001. **A**) Spatial autocorrelation of size classes determined by the distances between locations. **B**) All individuals included in each increased size class..... 82

3.7 Spatial autocorrelation 2009 samples. **A**) Spatial autocorrelation of size classes determined by the distances between locations. **B**) All individuals included in each increased size class..... 83

List of Appendices

2A Species included for the determination of sequence divergence between sparid genera based on mitochondrial DNA 16S rDNA sequences..... 56

2B Sequences obtained from 454 sequencing of microsatellite loci developed for (**A**) *P. blochii* and (**B**) *L. lithognathus*. Repeat sequences and primer binding sites are indicated in bold. 57

2C Homozygotes sequenced to assess repeat and flanking region variability of alleles at microsatellite loci of (**A**) *P. blochii* and (**B**) *L. lithognathus*, compared with the allele length from fragment analysis (shown in parentheses on the left of each sequence); DNA sequence and the number of repeats are provided for each sequences allele..... 59

2D Summary of cross-species for 19 microsatellite markers amplified in 11 sparid species. Amplification, size-range and percent null alleles (Oosterhout estimate) are recorded for each locus..... 61

2E Summary of the pairwise 16S sequence divergence (GD in %) and percentage predicted and observed amplification (AL in %) and polymorphism (PL in %) for the sparid genera with *P. blochii* and *L. lithognathus* as source species respectively..... 63

3A Summary statistics for each location in 2001 and 2009. The table includes number of alleles (N_A), Observed heterozygosity (H_O), Expected heterozygosity (H_E) and F_{IS}. Deviations from HWE are also indicated..... 92

3B Distance matrix between 2001 samples and 2009 samples..... 95

3C Contributions of individual markers to the SAC (not all data is included) for each year..... 96

Chapter 1

General Introduction

Abstract

The marine environment is expansive with a general lack of barriers restricting movement of marine organisms. Many marine fish species have adapted to this environment and exhibit large populations with only low levels of genetic subdivision. These species have developed life-history traits which allow for large populations with efficient dispersal potential (high fecundity, extended larval periods and adult migration). Despite their large population size and efficiency of dispersal, management and conservation of such species that are commercially harvested is still essential. These species are harvested in large quantities from their natural environments as a source of protein and this can be devastating to their survival. The reason is that often they have low effective population sizes (N_e), which implies that they essentially respond as a smaller population to evolutionary forces. To understand the genetic interactions of species within their environment, population genetic studies are required, and markers that are commonly used in these studies are microsatellites. These markers are very useful as they are co-dominant, tandem repeats displaying length polymorphism. Population structure ranges on a continuum from isolated populations on the one extreme to panmictic populations on the other. Identifying the level of structure of a species may contribute to the better understanding of its biology and ecology, which in turn will help the overall management of the species. South Africa has a diverse number of marine habitats with rich species diversity and many endemics. One of the main families of economic importance to the line fishery of South Africa is the Sparidae. This family comprises mostly reef or near shore demersal fish and 24 of the 110 species found around South Africa are endemics. Two endemic species of particular interest are the hottentot seabream (*Pachymetopon blochii*) and the white steenbras (*Lithognathus lithognathus*). These species are both considered vulnerable due to a combination of life-history traits and fishing pressure. This review focuses on the role of life-history traits and oceanographic features on restricting and/or enhancing gene flow of marine species and the implications for two southern African, economically important, sparid species.

Keywords: genetic variation, *Lithognathus lithognathus*, microsatellites, population genetics, *Pachymetopon blochii*, Sparidae

1.1 Introduction

By defining the degree of genetic spatial variation of populations, inferences can be made about what biological and environmental factors influence the genetic divergence and biogeography of marine species (Cowen *et al.* 2006). The marine environment is extensive, covering over 70 percent of the earth's surface and contain highly diverse habitats (Avise 1998; Chopelet *et al.* 2009). These habitats are not always continuous and may be patchy (e.g. kelp beds) and variable (Gratwicke & Speight 2005). The interactions of individuals of a species (life-history traits) within and between patchy habitats lead to a specific degree of genetic diversity and population structure that can be identified (Chopelet *et al.* 2009). Many marine species are often characterised by specific life-history traits such as high fecundity, pelagic eggs and larvae, high mobility and longevity, which contribute to high levels of gene flow counterbalancing spatial heterogeneity of the adults habitats (Avise 1998; Hauser & Carvalho 2008). These life-history traits coupled with the more homogenous habitats of larvae that have limited physical barriers (allowing dispersal between habitats) lead to a general lack of highly divergent populations at a regional scale. This is however not always the case for all organisms with several marine fish having been identified with population structure on local and regional scales (González *et al.* 2008; Narum *et al.* 2008; Nielsen *et al.* 2009). Population genetics plays a major role in the management and the conservation of commercial marine fish as the statistical approaches allow for the identification of the number of stocks that may need to be conserved separately. Marine fish are harvested in large quantities and this has led to overexploitation (Carvalho & Hauser 1998). By over-harvesting, genetic diversity may be lost, which has wider implications as this decreases the ability of the species to adapt (Fenberg & Roy 2008). Also marine fish are one of the last wild sources of protein and as the stocks become depleted it has far reaching implications on the people who rely on it as a food source and income (Ryman *et al.* 1995). The correct management and conservation required to maintain the sustainability of this resource starts with the improvement of knowledge of the population biology of the targeted species.

This review focuses on the impact of life-history traits and oceanographic features on the connectivity or genetic divergence of marine populations and the molecular techniques available to identify levels of population structure and gene flow. The southern African marine context is discussed in terms of the oceanographic features, line fisheries and coastal fishes. The main focus is on the demersal Sparidae of importance to the South African line fishery. Species biology is discussed for two vulnerable endemic sparid species.

1.2. Management and conservation in the marine realm

In the past, the marine species in the oceans were perceived as an endless resource. However, in the 1970's populations of some fish began to crash and fisheries yields declined (Devine *et al.* 2006). Management strategies were then implemented to control the number of tons caught annually but it was not deemed necessary to consider the resource management further than restricting commercial harvesting (Turpie *et al.* 2000). This, however, will not be enough to save and conserve species that are being exploited for commercial gain.

Marine species are the only organisms still harvested in large quantities from their natural environments for consumption by humans (Ryman *et al.* 1995; Carvalho & Hauser 1998; Hauser & Ward 1998). Identifying the population genetic structure through molecular markers provides another dimension to the management and the conservation of these exploited marine species (Avise 1998). This information should be used in combination with other analyses and parameters such as parasitology, tagging studies, species biology and oceanographic features to gain a complete strategy for effective management (Hauser & Ward 1998). Before effective management of a species can be implemented a thorough understanding of the species' biology, life-history traits, habitat, population structure and distribution is required (Cadrin *et al.* 2005). The main aim of management of marine fish populations (stocks) is to provide a balance between the needs of the fishermen and the restrictions placed to have a sustainable resource (Castilla 2000). Fish stocks are a shared resource between countries, which adds additional fishing pressure and complicates the management of these stocks further, as compensations need to be made due to social, political and economic pressure (Pawson & Jennings 1996; Castilla 2000).

When management of populations is not controlled and species are overharvested this has far-reaching effects which include changes in the ecosystems and biological changes in the exploited species as well as major declines in annual yields (Fenberg & Roy 2008). Size selectivity also places pressure on fish populations when restrictions only allow for the larger individuals to be caught. This leads to a negative impact on the demography of fish populations as the populations then mainly consist of juveniles and young adults that have not reached sexual maturity (Grant & Bowen 1998; Fenberg & Roy 2008). Such pressures will have major consequences for the population biology and in turn for the resource, since reduction in abundance and fecundity of adults will lead to smaller yields (Rochet 2000; Kuparinen & Merila 2007).

It is estimated by the Food and Agricultural Organisation (FAO) of the UN that as much as 52 % of current fisheries are fully exploited and 19 % are being overexploited (Garcia & Rosenberg 2010). This overexploitation may also lead to commercial or total extinction in three to four years if there is constant recruitment failure (Grant & Bowen 1998). These species of economic importance are harvested in quantities that are too large for the populations to recover adequately. In South Africa an example of this is the seventyfour seabream (*Polysteganus undulosus*). Adults were easily targeted through aggregations that formed during their spawning period (late winter) around the coast of KwaZulu-Natal until the 1980's; in 1998 a moratorium was placed on the exploitation of the species and it is now considered commercially extinct (Chale-Matsau *et al.* 2001).

Not all management strategies are inadequate and there are some fisheries which are considered sustainable. For instance, the Pacific Halibut (*Hippoglossus stenolepis*) fishery has been going since 1892 and is controlled by the International Pacific Halibut Commission (IPHC; Castilla 2000). Another example of a sustainable fishery is the Australian western rock lobster (*Panulirus cygnus*) fishery. These two fisheries are managed according to a research-based strategy. The stocks are constantly monitored and catch limits are strictly implemented. In addition the management programs are constantly adapted to suit how the stocks are progressing and changing (Castilla 2000). There have also been records of populations recovering after population crashes. An example of this is the North sea cod (*Gadus morhua*) population at Flamborough Head that had two successive population declines between 1954 and 1998. This study indicated a decrease in genetic diversity from 1954 - 1970 and then an increase (recovery) from 1970 – 1998. However, it is thought that this population did not completely recover on its own, but rather, was aided by individuals that migrated from another population (Hutchinson *et al.* 2003).

Before adequate management can be implemented the interaction between the environment of the species of interest, the specific life-history traits and its species biology needs to be considered. These interactions are linked to the degree of spatial genetic variation identified.

1.3. Life-history traits and marine factors influencing population genetics in marine fish

Within the marine realm, habitats are extremely diverse and range from sandy beaches to deep ocean vents (Heemstra & Heemstra 2004). The diverse habitats are dependent on factors such as varying levels of salinity, temperature and pressure and these variations may act as barriers to the movement of species (Waples 1998). As the depth of the ocean floor

increases the temperature decreases and the pressure increases (Carvalho 1993). Marine fish are restricted by the conditions in the various habitats and often have varying life-history traits. These life-history traits in combination with the level of patchiness of habitats may lead to various spatial models of population structure, from subdivision into divergent populations at the one extreme to the existence of one panmictic population at the other extreme (Fig. 1.1; Waples & Gaggiotti 2006).

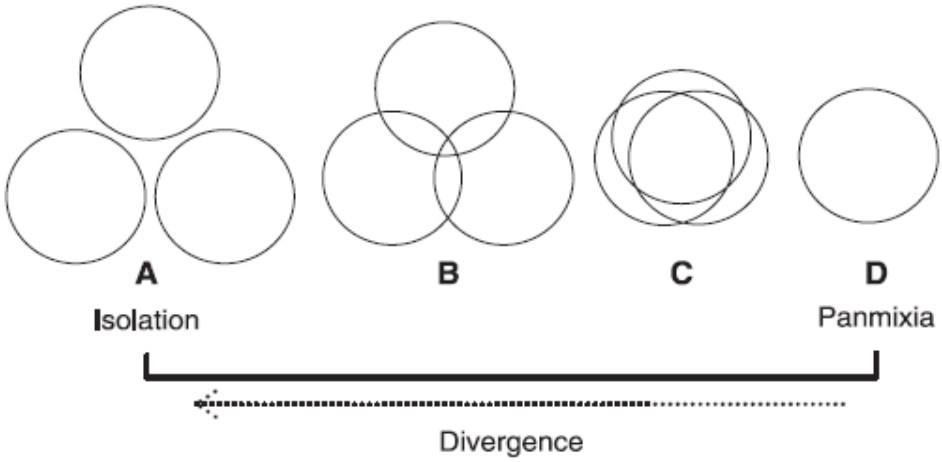


Fig. 1.1 Population models differing by their level of population divergence from completely isolated sub-populations (A) to panmixia (D). A) Three isolated sub-populations, B) three sub-populations with low connectivity C) sub-populations with high connectivity and D) panmixia. Diagram from Waples & Gaggiotti (2006).

Over time natural populations may change in size as well as density and location (Hey & Machado 2003). The populations are constantly changing and can divide into sub-populations or join with other populations. Researchers attempt to test whether their data fit one of various theoretical population structure models. Some examples are the island model of migration, the stepping-stone model, isolation by distance and the metapopulation model (Hey & Machado 2003). In marine organisms, populations have long been thought to follow the panmixia model (random mating, infinite population size; Hauser & Carvalho 2008).

Marine fish are generally widely dispersed, with strong dispersal ability and large population sizes throughout the ocean. Apart from the pelagic species, most marine fish species have their life divided into two stages: A relatively sedentary adult stage often specialized to a specific habitat and a planktonic propagule stage (eggs and larvae; Van der Elst 1988). It is during this latter stage that eggs and larvae passively move with the ocean currents (Cadrin *et al.* 2005). This means that the offspring of such species encounter unpredictable

environments and many of them are lost before they reach maturity and recruit into the adult populations. To overcome these lost, many species develop a strategy to produce as many offspring as possible so that some may survive. In consequence, the high gene flow potential in those species are associated with the numerous propagules and their dispersal ability (Cowen *et al.* 2006), but also coupled with moderate migration of adults and the large population sizes in addition to limited physical barriers. All these factors limit the divergence between different marine populations (Ward 2000; Cadrin *et al.* 2005; Zhan *et al.* 2009).

Barriers that restrict gene flow in the marine environment may be physical, such as some of the factors mentioned above (i.e. salinity), as well as upwelling zones, currents and eddies (Hauser & Carvalho 2008). Additionally, they may also be historical and may not be considered barriers anymore. Past glaciations and sea level changes lead to historical barriers. An example is the various ice ages which led to sea level changes and the formation of land bridges restricting the movement of marine species between the Atlantic and Mediterranean Sea (Patarnello *et al.* 2007).

The marine environment comprises of ecological zones. The open ocean (pelagic zone) is divided into different layers: The upper pelagic from 0 - 200 m in depth; the mesopelagic zone (200 - 1000 m) and the bathypelagic zone (from 1000 - 4000 m; Heemstra & Heemstra 2004). In these zones, pelagic fish are found which have traits such as extensive mobility with annual migrations (Hauser & Ward 1998). The benthic (bottom) zone along the ocean floor includes from the deepest areas to the shallow waters of the coastal areas (Heemstra & Heemstra 2004). Marine biodiversity is higher in the shallow-water benthic zone due to the increased diversity in habitats (Van der Elst 1988; Gratwicke & Speight 2005). This zone also has a patchiness of habitats which may lead to population subdivision. Demersal (living near the bottom) and benthic (living on the bottom) species often have restricted movement as the adults cannot cross deeper water bodies. These species are also less active and may have localised distributions where the habitat is suitable (Van der Elst 1988).

Dispersal ability of the adults and pelagic eggs and larvae, are thought to substantially contribute to the low genetic divergence identified in marine fish populations (Hauser & Carvalho 2008). Marine fish have a high fecundity and each female may release hundreds of thousands of eggs during spawning events. These spawning events may occur annually or multiple times every year. The pelagic propagules are transported passively by the oceanographic features (currents, eddies, upwelling zones) which may lead to high levels of mixing and dispersal of the larvae over many kilometres (Shanks & Eckert 2005; O'Connor *et al.* 2007). Larval duration has a link with the ocean temperature where a trend has been

detected with longer larval duration and cooler ocean temperature (O'Connor *et al.* 2007). Propagules may therefore be dispersed further through currents during a longer larval duration in temperate marine environments. Also this may be coupled to higher connectivity observed at the higher latitudes (towards the temperate environments), where the oceans are substantially cooler than near the tropics (lower latitudes; Kelly & Eernisse 2007). However, a recent review across 87 studies on marine larval dispersal and population structure covering 130 species identified no clear correlation between population structure and larval dispersal, which implies that it is probably a combination of factors that leads to structuring in the marine environment (Weersing & Toonen 2009).

Several of the commercially exploited marine species have life-history traits that make them vulnerable to overexploitation if fished in large numbers each year. These traits are sedentary adults, late sexual maturity, low fecundity and a restricted distribution range (Cadrin *et al.* 2005). Also, it has been found that although marine fish have large population sizes, their effective population size (N_e) may be low due to the variation in reproductive success. This implies that although the large population sizes influence demographic interactions, on an evolutionary scale the populations are small and may be more strongly affected by evolutionary forces (genetic drift, mutations, selection and migration; Hauser & Carvalho 2008; Waples 2010).

Marine fish species that have already shown huge reductions in population sizes in recent years are Atlantic cod (*Gadus morhua*), tuna (*Thunnus alalunga*) and swordfish (*Xiphias gladius*; Devine *et al.* 2006). Species that are exploited commercially need to be studied so that accurate and efficient management may be implemented. Through an extrapolation of diversity depletion of marine fish and invertebrate species, Worm *et al*. (2006) estimated that several fisheries resources will be totally depleted by 2048, however, this has been criticized by Longhurst (2007) who pointed out a number of inconsistencies in data assimilation and interpretation.

Molecular markers have become a common tool to help with identifying underlying population structure (O'Connell & Wright 1997). Patterns of molecular variance identified by molecular markers on spatial or temporal scales for a species may be used to infer population structure.

1.4. Molecular markers

Through the development of molecular markers it has become easier to study species in the marine environment as its inaccessibility has hindered marine studies in the past. These markers allow inferences to be made about the population without having to directly study the species in its environment (Grant & Bowen 1998; Liu & Cordes 2004). These markers have provided insight into genetic variability, species identification, inbreeding, traceability and population structure (Liu & Cordes 2004). The choice of the most applicable marker to a study will depend on the question being asked (Freeland 2005).

When population divergence was first studied, markers such as allozymes, amplified fragment length polymorphisms (AFLPs), random amplified polymorphic DNAs (RAPDs), and restriction fragment length polymorphisms (RFLPs) were the markers of choice (Liu & Cordes 2004). Since then there has been a focus on mitochondrial DNA (mtDNA). This is because it can easily be amplified across many species, is present in multiple copies in most cells, is generally haploid and has a high rate of mutation. However, mtDNA is only inherited from the maternal parent and the whole mtDNA genome needs to be considered as a single linked locus (Liu & Cordes 2004). In a case where differentiation of closely related individuals is required, markers that have a higher variability are needed. The markers that are considered for population studies are generally co-dominant as they provide all the information for both alleles in diploid individuals (Freeland 2005). Nuclear markers which are currently available in population genetics studies are microsatellites, SNPs (single nucleotide polymorphisms; Liu & Cordes 2004) and EPIC (exon-primed intron-crossing) markers (Chenuil *et al.* 2010).

Although microsatellites are commonly used as the co-dominant marker of choice, EPIC markers have some advantages over microsatellites as they have lower levels of null alleles and are applicable across many more species (Li *et al.* 2010). Here, however, focus is placed on microsatellites as these markers are species-specific, are characterized by high variability and are relatively time and cost effective once the markers are developed (Zane *et al.* 2002). The development of microsatellite markers has also been made increasingly easier through the wider availability of high-throughput sequencers such as the 454 LifeSciences/Roche GS-FLX (pyrosequencing; Hudson 2007; Santana *et al.* 2009). At present microsatellite markers are the most common markers used in population genetics studies and are particularly useful in identifying population structure (Ellegren 2004; Waples & Gaggiotti 2006).

1.4.1 Microsatellites

Microsatellites are tandem repeats with varying repeat classes from di- to pentanucleotide repeats found in both bacteria and eukaryotes (Hansen *et al.* 2001; Zane *et al.* 2002; Cadrin *et al.* 2005). These repeats are mainly identified in non-coding regions. The polymorphism of these molecular markers is measured through the varying lengths of alleles at specific loci (Ellegren 2004; Cadrin *et al.* 2005). The repeats that are most common in marine fish are $(AC)_n$, $(ACC)_n$ and $(GATA)_n$ and these can be from 5-100 tandem repeats (Cadrin *et al.* 2005).

The abundance of microsatellites has been estimated from 10^3 - 10^5 throughout eukaryote genomes (Cadrin *et al.* 2005). Fish genomes contain a high number of microsatellites, estimated to occur every 1500 kb (O'Connell & Wright 1997). The estimated mutation rate for microsatellites per generation is 10^{-2} - 10^{-6}, giving these DNA fragments a relatively high rate of mutation (Chistiakov *et al.* 2006). The repeat length polymorphism seen in these microsatellites is thought to either occur through the slippage of the DNA polymerase during the replication process or through unequal recombination events (Tóth *et al.* 2000).

Several mutation models are proposed for microsatellite loci (Bhargava & Fuentes 2010), of which three are commonly applied. The first model is the stepwise mutation model (SMM), this is where the microsatellite either gains or loses a single repeat at a time. The second model is the infinite allele model (IAM) and this is where each new mutation event occurring at a locus gives rise to a new allelic state not previously observed (Zane *et al.* 2002; Cadrin *et al.* 2005). The third model known as the two-phase model (TPM) is a combination of the stepwise mutation model (SMM) and the K allele model and is likely to represent a more realistic mutation model for microsatellites (Di Rienzo *et al.* 1994; Cadrin *et al.* 2005).

Microsatellites are relatively easy to isolate especially in genomes that are rich in microsatellites (Zane *et al.* 2002), for example marine fish. One of the most effective microsatellite isolation procedures is the Fast Isolation by AFLP of Sequences COntaining Repeats (FIASCO) developed by *Zane et al.* (2002). This procedure involves a digestion/ligation step, probing and enrichment for microsatellite repeats (Zane *et al.* 2002). Tetranucleotide repeat probes are often chosen over di- or trinucleotides as there is less slippage during polymerase chain reaction amplification which leads to more accurate genotypic scoring (Kruglyak *et al.* 1998).There are also potential disadvantages to using tetranucleotide repeats, such as higher possibilities of allelic drop out (Broquet *et al.* 2007) This may be due to the extensive size differences between alleles in some tetranucleotides

that have many repeats. It has been shown that short alleles amplify more successfully during the PCR and the larger allele is outcompeted, leading to it not being detected (Wattier *et al.* 1998). This, however, can be controlled by excluding microsatellites that show an excess of homozygotes. These repeat fragments are either isolated through cloning or more recently these fragments have been successfully sequenced through pyrosequencing™ (Allentoft *et al.* 2009; Santana *et al.* 2009). From this, primers are designed from the flanking regions of the identified repeats.

Key applications of microsatellite markers in marine studies are summarized here, these include: Analyzing and determining population structure, identify different stocks and assigning individuals to specific populations of a species (Selkoe & Toonen 2006). Microsatellites provide insight into gene flow between populations and within a population. These markers can identify recent bottlenecks, effective population size, migration rates and the relatedness between individuals (parentage assignment; Hansen *et al.* 2001). Amplification of microsatellites requires minimal amounts of DNA which can be sampled from tissue invasively or non-invasively (Selkoe & Toonen 2006). In terms of marine fish this allows for non-invasive sampling where just a small part of the fin is required and the fish may then be released back into the ocean.

No molecular marker is perfect and there are some disadvantages to using microsatellite markers. Homoplasy, null alleles, genotyping errors and complicated mutation models that are not fully understood are common disadvantages (Narum *et al.* 2008). Homoplasy is a major disadvantage as alleles may be the same size by state but not by decent. However, if the level of homoplasy is relatively low it does not influence the study substantially. Some forms of homoplasy can be controlled for by sequencing homozygotes and confirming the repeat length and absence of mutation in the flanking regions. If high levels of homoplasy are detected the marker should be removed from further analyses.

Allelic dropout can also occur when only one allele amplifies and the individual's genotype is mistakenly thought to be homozygous (Morin 2004). When mutations occur in the flanking regions of the microsatellites, some alleles may consistently not amplify (known as null alleles) or may appear as an incorrect fragment length (Pompanon *et al.* 2005). Null alleles may be identified through summary statistics or analysis programs and are indicated by the excess of homozygotes. If these levels are sufficiently high the marker should be removed from the study (Wagner *et al.* 2006). There are protocols that can be followed to identify and correct for these genotyping errors (Pompanon *et al.* 2005). If the DNA quality or quantity is very low this may also lead to allelic dropout. Many of these errors may be eliminated by

amplifying a locus more than once and estimating error rate per locus (Bonin *et al.* 2004; Pompanon *et al.* 2005).

1.4.2 Cross-species amplification

Microsatellite markers, which are generally developed for a specific species of interest, may also be tested in closely related species for amplification success and levels of polymorphism. This saves the time and money of developing a *de novo* set of microsatellites (Carreras-Carbonell *et al.* 2008). This is possible through the relatively conserved flanking regions so that the same microsatellite loci may be amplified in different, closely related species (Zane *et al.* 2002). Extensive cross-species amplification of microsatellite loci has been tested in the Sparidae (Brown *et al.* 2005; Chen *et al.* 2005; Hatanaka & Yamada 2006; Jean *et al.* 2006; Castro *et al.* 2007; Roques *et al.* 2007a; Roques *et al.* 2007b; Navarro *et al.* 2008).

Although there are benefits to identifying microsatellites in this manner, often many markers will need to be tested before an adequate set of markers are found (Liu *et al.* 2007). Through cross-species amplification there is the potential for increased null alleles (Bonin *et al.* 2004; Soulsbury *et al.* 2009) due to potential mutations occurring in the primer binding sites where the original primers were designed. This can often be rectified by sequencing a few homozygotes and re-designing the primers (Soulsbury *et al.* 2009).

A question often asked in terms of cross-species amplification is to what extent an isolated microsatellite locus could be polymorphic in other related species. The level of conservation may vary, with one marine fish microsatellite locus being conserved over 470 million years of divergence (Rico *et al.* 1996). Many studies have identified a strong negative correlation between the genetic distance of the species of interest from the source species (species from which microsatellites were developed) and polymorphism. This means that the more genetically distant the species is from the source species, the less likely it is for polymorphic loci to be identified. This negative correlation is found across a diverse number of taxa (Primmer *et al.* 2005; Barbara *et al.* 2007; Carreras-Carbonell *et al.* 2008). Carrera-Carbonell and colleagues identified a negative correlation between genetic distance determined from mitochondrial genes (12S & 16S rDNA) and the total number of loci that amplify and are polymorphic in marine fish. This information can be used to predict which marine fish species is the most likely to produce polymorphic loci in a species of interest before testing of loci begins (Carreras-Carbonell *et al.* 2008).

Before the results for microsatellite markers can be interpreted on a population genetic level, information is required about the distribution, species biology and fishing pressures of the species being studied. South Africa has a diverse coastline maintaining varied species and high fishing pressure, with many species currently in need of scientific study and management.

1.5. South African marine context

The rich biodiversity along the South African coast may largely be due to the diversity in habitats that are available. These habitats include deep oceans, coral reefs, sandy beaches, estuaries and rocky coasts (Heemstra & Heemstra 2004). Around southern Africa there are three main biogeographical regions: (1) The cool-temperate which ranges from around Cape Point into Namibia and has cold, nutrient rich water (12 - 13 $^{\circ}$C) from the Benguela Current (Anderson *et al.* 2009), (2) the warm-temperate region along the south coast (Cape Point to the former Transkei); and (3) the subtropical region along the northeast coast with sea temperatures between 17 – 18 $^{\circ}$C (Fig. 1.2; Anderson *et al.* 2009). Variation in temperature, salinity and other physical factors can be attributed to the the warm Agulhas Current and cold Benguela Current (Heemstra & Heemstra 2004; Fig. 1.2).

Species richness increases from the west coast to the east coast of South Africa (Turpie *et al.* 2000). The Benguela Current flows northwards and extends from Cape Agulhas (35°S) to the Angolan port of Namibe (15°S) (Sakko 1998). The characteristics of this eastern boundary current upwelling system (Benguela Current) is the cold surface water and increased biological activity; this does not imply an increased number of species but rather an increased abundance per species due to the higher levels of nutrients found in these waters (Sakko 1998; Nicholson 2010). The Benguela Current consists of Indian and South Atlantic subtropical thermocline water (Turpie *et al.* 2000). The upwelling around Lüderitz causes unstable conditions in this area (Sakko 1998).

Warm tropical water brought by the Agulhas current, a powerful western boundary current, flows along the east coast of South Africa. This current flows in a southerly direction and provides a more stable habitat, which is rich in species from the tropical Indo-Pacific and Western Indian Ocean (Turpie *et al.* 2000; Lutjeharms *et al.* 2001).

Around South Africa three factors determine the distribution of species along the coast. These are biological interactions, temperature and the geology of the area (Turpie *et al.* 2000). The currents found along the coastline of southern Africa are the most influential

factors as they control the temperature of the sea water and the environmental productivity. Larval and adult dispersal relies on the movement of these ocean currents (Turpie *et al.* 2000).

Focusing on the west coast, a system for dispersing larvae in small pelagic species (e.g. sardine and anchovy) is known, where species that spawn around the Agulhas Bank have their pelagic propagules (eggs & larvae) transported by an inshore "jet" current moving in a northerly direction (black arrows; Fig. 1.2). This current feeds into inshore currents (red arrow) moving in a southerly direction along the coast and depositing the propagules into suitable habitats such as estuaries (Hutchings *et al.* 2009). The main west coast oceanographic features are indicated in Fig. 1.2.

1.5.1 Line fisheries around South Africa

The South African line fishery harvests more than 200 species of marine fishes and is controlled through bag limits and catch quotas (Griffiths *et al.* 1999). The Western Cape and Eastern Cape Provinces' inshore fisheries are made up of the net and line fishery. This line fishery includes angling and commercial boat-based fishing. The predominant species caught through the line fishery are several sparids and snoek (*Thyrsites atun*; Turpie *et al.* 2003).

The Cape commercial line fishery operates around the western coast from the Orange River to the Kei River from 5 - 130 m in depth following the continental shelf (Griffiths 2000). Forty species are targeted along the Cape coastline and of these 20 are of economic importance. In 1997 the reported annual catch by the line fishery was approximately 15 500 tons annually and this is 95 % of the total line fishery catch for South Africa (Griffiths 2000). Therefore, the majority of the line fish are caught along the Cape coastline and only five percent of the annual commercial line fishery catch comes from the eastern coast i.e. from Kei River to Mozambique. The Western Cape Province relies on fisheries to provide a large portion of its annual income. The line fisheries also support other industries such as equipment, boating and providers of bait (Turpie *et al.* 2003).

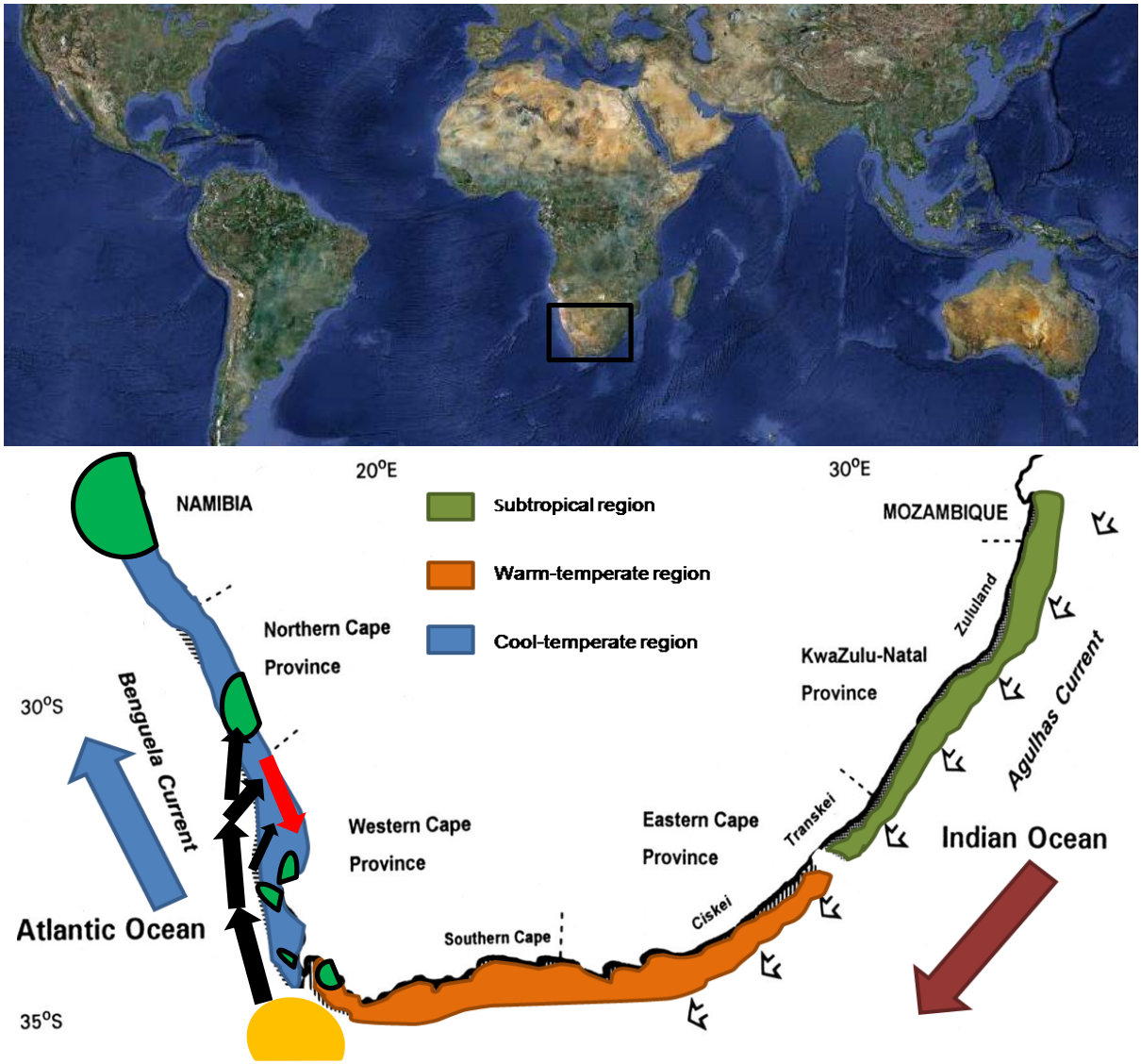


Fig. 1.2 Map of the world (google maps) with a focus on South Africa. The enlarged map shows the biogeographical provinces and currents around South Africa. Spawning area (yellow), upwelling zones (green) are shown on the west coast. The "jet" current (black arrows) moves along the west coast and feeds into an inshore current (red arrow). Modified from Whitfield *et al*. (2001).

1.6. Coastal fishes of southern Africa

1.6.1 The Fish fauna

Southern African fish fauna makes up 16% of the total marine fish species in the world (Heemstra & Heemstra 2004) with about 2200 species identified from this region. Approximately 286 of these species are endemic to southern Africa (Van der Elst 1988). A total of 1800 species are shallow-water coastal species and these represent 200 families, which is 80 % of the world's shallow-water families (Heemstra & Heemstra 2004). Two

hundred and twenty seven shallow-water species around southern Africa are endemics (Turpie *et al.* 2000), among which the most prevalent families are the Clinidae (klipfishes), Sparidae (seabreams) and Gobidae (gobies). The majority of these species are found in shallow-waters with only a few residing at depths of more than 200 m. A general characteristic is that these species also tend to have small home ranges (sedentary) and distributions which makes them more vulnerable to overexploitation and habitat loss (Turpie *et al.* 2000).

1.6.2 The family Sparidae

Across the globe the sparids are of major economic significance (Van der Elst 1988; De La Herran *et al.* 2001; Leis *et al.* 2002; Franch *et al.* 2006). This family has a broad geographical distribution with the majority of species found in the Indian Ocean (Turpie *et al.* 2000). Fishes from this family are also dispersed through the tropical and temperate Pacific and Atlantic oceans and estuaries (Leis *et al.* 2002; Orrell & Carpenter 2004; Sheaves 2006). These species' habitats vary in depth (0 - 250 m) and they are mainly shallow-water species (Bargelloni *et al.* 2003). Of the 110 sparid species around South Africa 24 are endemic (Van der Elst 1988; Heemstra & Heemstra 2004). Only a few sparids have a distribution on the western coast of South Africa, namely the hottentot seabream (*Pachymetopon blochii*), west coast steenbras (*Lithognathus aureti*), white steenbras (*Lithognathus lithognathus*) and white stumpnose (*Rhabdosargus globiceps*; Van der Elst 1988; Heemstra & Heemstra 2004). Some sparids are slow-growing, reach sexual maturity late, undergo sex changes and have small home ranges. These life-history traits place these species at risk of overexploitation. Sparids already placed under immense pressure through overharvesting include the white steenbras which is considered a lower risk species that is conservation-dependent, the red porgy (*Pagrus pagrus*) that is endangered (www.IUCNRedList.org) and the seventyfour seabream (*Polysteganus undulosus*) which is commercially extinct (Chale-Matsau *et al.* 2001).

For Sparidae numerous microsatellite loci have been developed including markers identified through cross-species amplification (Brown *et al.* 2005; Chen *et al.* 2005; Hatanaka & Yamada 2006; Jean *et al.* 2006; Piñera *et al.* 2006; Ponce *et al.* 2006; Castro *et al.* 2007; Piñera *et al.* 2007; Roques *et al.* 2007a; Roques *et al.* 2007b; Pérez *et al.* 2008). SNPs have been developed for the gilthead seabream *Sparus aurata* (Cenadelli *et al.* 2007) with a linkage map (Franch *et al.* 2006). Also there are two EST libraries available, one for *S. aurata* with 50 930 sequences on NCBI and *Lithognathus mormyrus* with 8109 sequences. Population structure has been determined for some commercially important sparids. These

have varied from weak spatial and temporal variation to panmixia (Planes & Lenfant 2002; Burridge & Versace 2007; González-Wangüemert *et al.* 2007; González-Wangüemert *et al.* 2010; Roberts & Ayre 2010; Teske *et al.* 2010).

A sparid phylogeny was determined by Orrell & Carpenter (2004) and then more recently by Chiba *et al.* (2009). Initially sub-families were identified through dentition: Boopsinae (BO), Denticinae (DE), Pagellinae (PE), Sparinae (SP) Diplodinae (DI) and Pagrinae (PA; De La Herran *et al.* 2001; Orrell & Carpenter 2004). Using partial mitochondrial 16S rRNA and cytochrome *b* sequences, Orrell & Carpenter (2004) found that their phylogeny validated six subfamilies but they were not all monophyletic; they also identified two distinct clades within the Sparidae. A more extensive phylogeny including more species was then completed in 2009, using cyt *b*, which also identified two separate clades (Chiba *et al.* 2009).

The present study focuses on two endemic sparid species from South Africa, namely the white steenbras (*Lithognathus lithognathus*) of the Pagellinae subfamily and the hottentot seabream (*Pachymetopon blochii*) of the Boopsinae subfamily. Both these species are grouped into clade B.

1.6.3 Lithognathus lithognathus

The white steenbras' distribution is believed to be between the Tugela River and Angola with a core distribution range from the Transkei to False Bay (Fig. 1.3). Some individuals are found on the west and south coast but these are relatively rare (Heemstra & Heemstra 2004). This species feeds mainly on invertebrates and occurs to a depth of 25 m. The adults are thought to annually migrate from the Western Cape to spawn off the northern Eastern Cape during June to August (Bennett 1993a; Heemstra & Heemstra 2004). The juveniles spend their first year in estuaries before moving to the surf zone. This dependence on estuaries has made this species particularly vulnerable to habitat degradation and easy targets for fishermen (Bennett 1993a). In recent years the catches of white steenbras have significantly declined due to degradation of the estuaries and increasing fishing pressure (Bennett 1993b; Heemstra & Heemstra 2004).

1.6.4 Pachymetopon blochii

Pachymetopon blochii, commonly known as the hottentot seabream, is distributed along the west coast of southern Africa from Angola to Cape Point in the south (it is rarely found past this point; Van der Elst 1988; Punt *et al.* 1996). It has been noted that it is very rarely found

past the upwelling zone around Lüderitz (Namibia), thus it's main and most prevalent distribution is along the west coast of South Africa (Fig. 1.3). The hottentot is considered vulnerable as individuals are believed to have a small home range, thus being relatively sedentary (more easily targeted by fishermen), and only sexually mature around five years (22 cm; Pulfrich & Griffiths 1988b; Heemstra & Heemstra 2004). This species is a serial spawner with spawning occurring year round, with a slight increase during late autumn and summer (Pulfrich & Griffiths 1988b). The females release on average 120 000 eggs each (Pulfrich & Griffiths 1988b).

This species is an omnivore and feeds on seaweed, molluscs and other crustaceans; as the fish becomes older its diet mainly consists of algae that grows within the kelp beds (*Ecklonia maxima* and *Laminaria pallida*). The habitats they can be found in are kelp beds and rocky cliffs in the ocean around a depth of 50 m. The juveniles are found in the same habitats as the adults so it is thought that spawning occurs within the adult habitats (Pulfrich & Griffiths 1988a). Unlike many species of Sparidae the hottentot seabream does not change sex and has a lifespan of approximately 12 years (Van der Elst 1988; Heemstra & Heemstra 2004).

Hottentot are caught through line fishing and small fishing boats but also often by larger fishing boats as bycatch. In 1988 a study by Pulfrich & Griffiths determined that this species was fully exploited and there are restrictions in place with a bag limit of ten individuals of a minimum length of 22 cm (Pulfrich & Griffiths 1988a). Although this species does not play a large economic role in the country compared to the larger deep sea catches, it forms part of the hand line fishery that makes up five percent of the South African Fisheries annual catch (Pulfrich & Griffiths 1988a). This species is important to local fishermen and is sold at local markets to sustain these fishing families. The hottentot is an important source of protein to some of the rural communities along the west coast of South Africa (Van der Elst 1988).

Along the western coast the two main marine fish species that are caught by line fishermen are the hottentot seabream (~ 600 tons annually) and snoek (6000 tons annually; Pulfrich & Griffiths 1988a; Sauer *et al.* 2003). The hottentot catches increase when the snoek fishery is closed and the species is therefore not always targeted (Pulfrich & Griffiths 1988a). There has not been a need to place too many restrictions on this species because it spawns throughout the year and when there is an increase in spawning in late autumn and summer, other fisheries are in favor leading to fewer *P. blochii* being caught (Pulfrich & Griffiths 1988a). During the hottentot's major spawning season, snoek and rocklobster are in demand and this in effect protects the hottentot. The *P. blochii* line fishery season is from April to October. The highest catches of this species are in Gordons Bay and Kalk Bay (Pulfrich &

Griffiths 1988a). As this species is non-migratory it has been deduced that if management is required marine protected areas (MPAs) would be beneficial (Pulfrich & Griffiths 1988a)

Fig. 1.3 Distribution range of the hottentot seabream (red) and the white steenbras (yellow). The dashed lines indicate where each species is thought to occur infrequently; the solid lines indicate the main distribution of the species.

1.7. Aim and Objectives

The main aim of this research project was to identify the spatial and temporal variation within the hottentot seabream along the west coast of South Africa using microsatellite markers. This was achieved by developing microsatellite markers for the hottentot seabream and white steenbras and using these markers, as well as some loci developed in other sparids, as a panel to identify as many applicable markers as possible in the hottentot seabream. These markers were then genotyped in a number of hottentot individuals and different summary statistics and analyses performed to determine levels of gene flow and population differentiation in this species along the west coast of South Africa.

To achieve this, the following research objectives were addressed:

Chapter 2

1. Identifying a sufficient set of microsatellite markers for the hottentot seabream and white steenbras through *de novo* development and cross-species testing.
2. Testing the applicability of these markers for other economically important sparids around South Africa.
3. Identification of a relationship between genetic distance and microsatellite amplification and polymorphism between closely related sparid species.
4. Evaluating the usefulness of 454 sequencing as an alternative to cloning in the process of developing microsatellite markers in marine fish.

Chapter 3

1. Inferring the number of populations of the hottentot seabream along the west coast of South Africa
2. Evaluating spatial and temporal genetic variations of the species using spatial (across distribution range) and temporal (2001 and 2009) sampling surveys.

1.8. Dissertation outline

Chapter 2: Isolation and characterization of microsatellites in two South African endemic sparids *Pachymetopon blochii* (Valenciennes 1830) and *Lithognathus lithognathus* (Curvier 1829) and cross-species amplification in nine other economically important sparids

The first section of this study focused on the development of two separate species-specific microsatellite marker sets, one for the white steenbras (*Lithognathus lithognathus*) and the other for the hottentot seabream (*Pachymetopon blochii*) using the FIASCO protocol and 454 LifeSciences/Roche GS-FLX (pyrosequencing™). These developed microsatellites, as well as markers identified through cross-species testing, were used as a panel to investigate cross-species amplification in eleven economically important sparids (nine other sparid species and the hottentot and white steenbras). Nine microsatellites were identified as polymorphic in each focal species with relatively high levels of polymorphism. Fifteen of these newly developed loci and four loci identified from other sparids were included as a panel of markers on the eleven species of sparids. A number of loci were identified for each species included in the study. In total, 14 suitable loci were identified as polymorphic in the hottentot seabream and 12 in the white steenbras. The markers identified as polymorphic in the white steenbras are being used in a study by a PhD student at SAIAB (The South African Institute for Aquatic Biodiversity) that combines information from population genetics and telemetry to understand the connectivity and movement patterns of this species. The hottentot seabream study focused on identifying the spatial and genetic variation around South Africa (see Chapter 3).

This chapter will be submitted to *Molecular Ecology Resources*

Chapter 3: Genetic differentiation in the hottentot seabream *Pachymetopon blochii* (Valenciennes, 1830) from South Africa based on microsatellite variation

The main focus of the present study was on identifying the spatial and temporal genetic variation in the hottentot seabream along the west coast of South Africa. This species was of interest as it is the second most targeted line fish species on the west coast of South Africa after snoek (*Thyrsites atun*). It also has life-history traits (sedentary adults and late sexual maturity) which make it more susceptible to overexploitation. This study identified only weak spatial and temporal variation. The observed population structure on temporal and spatial

scales is thought to be due to variation in reproductive success which leads to random genetic patchiness.

This chapter will be submitted to *Marine Ecology Progress Series*

Chapter 4: Dissertation synthesis

This chapter provides a synthesis of the results from Chapter 2 and 3.

General notes

The authors that will be involved in the publication of any articles from this research will include Kerry Reid, Thierry B. Hoareau and Paulette Bloomer. This dissertation has followed the referencing style of *Molecular Ecology*. Reference lists are provided for each chapter thus some reference duplication was unavoidable.

1.9. References

Allentoft ME, Schuster SC, Holdaway RN*, et al.* (2009) Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. *BioTechniques* **46**, 195-200.

Anderson RJ, Bolton JJ, Stegenga H (2009) Using the biogeographical distribution and diversity of seaweed species to test the efficacy of marine protected areas in the warm-temperate Agulhas Marine Province, South Africa. *Biodiversity Research* **15**, 1017-1027.

Avise JC (1998) Conservation Genetics in the Marine Realm. *Journal of Heredity* **89**, 377-382.

Barbara T, Palma-Silva C, Paggi GM*, et al.* (2007) Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology* **16**, 3759-3767.

Bhargava A, Fuentes FF (2010) Mutational dynamics of microsatellites. *Molecular Biotechnology* **44**, 250-266

Bargelloni L, Alarcon JA, Alvarez MC*, et al.* (2003) Discord in the family Sparidae (Teleostei): divergent phylogeographical patterns across the Atlantic-Mediterranean divide. *Journal of Evolutionary Biology* **16**, 1149-1158.

Bennett BA (1993a) Aspects of the biology and life history of white steenbras *Lithognathus lithognathus* in southern Africa. *South African Journal of Marine Sciences* **13**, 83-96.

Bennett BA (1993b) The fishery for white steenbras *Lithognathus lithognathus* off the cape coast, South Africa, with some considerations for its management. *South African Journal of Marine Sciences* **13**, 1-14.

Bonin A, Bellemain E, Bronken Eidesen P*, et al.* (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* **13**, 3261-3273.

Brown RC, Tsalavouta M, Terzoglou V, Magoulas A, McAndrew BJ (2005) Additional microsatellites for *Sparus aurata* and cross-species amplification within the Sparidae family. *Molecular Ecology Notes* **5**, 605-607.

Burridge CP, Versace VL (2007) Population genetic structuring in Acanthopagrus butcheri (Pisces: Sparidae): Does low gene flow among estuaries apply to both sexes? *Marine Biotechnology* **9**, 33-44.

Cadrin SX, Friedland KD, Waldman JR (2005) *Stock identification methods: Applications in fishery science* Elsevier Academic Press. Amsterdam.

Carreras-Carbonell J, Macpherson E, Pascual M (2008) Utility of pairwise mtDNA genetic distances for predicting cross-species microsatellite amplification and polymorphism success in fishes. *Conservation Genetics* **9**, 181-190.

Carvalho GR (1993) Evolutionary aspects of fish distribution: genetic variability and adaptation. *Journal of Fish Biology* **43** (Supplement A), 53-73.

Carvalho GR, Hauser L (1998) Advances in molecular analysis of fish population structure. *Italian Journal of Zoology* **65**, 21-33.

Castilla JC (2000) Roles of experimental marine ecology in coastal management and conservation. *Journal of Experimental marine Biology and Ecology* **250**, 3-21.

Castro J, Pino A, Hermida M*, et al.* (2007) A microsatellite marker tool for parentage assessment in gilthead seabream (*Sparus aurata*). *Aquaculture* **272**, 210-216.

Cenadelli S, Maran V, Bongioni G*, et al.* (2007) Identification of nuclear SNPs in gilthead seabream. *Journal of Fish Biology* **70**, 399-405.

Chale-Matsau JR, Govender A, Beckley LE (2001) Age, growth and retrospective stock assesment of an economically extinct sparid, *Polysteganus undulosus*, from South Africa. *Fisheries Research* **51**, 87-92.

Chen SL, Liu YG, Xu MY, Li J (2005) Isolation and characterization of polymorphic microsatellite loci from an EST-library of red sea bream (*Chrysophrys major*) and cross-species amplification. *Molecular Ecology notes* **5**, 215-217.

Chenuil A, Hoareau TB, Egea E*, et al.* (2010) An efficient method to find potentially universal population genetic markers, applied to metazoans. *BMC Evolutionary Biology* **10**, 1-17.

Chiba SN, Iwatsuki Y, Yoshino T, Hanzawa N (2009) Comprehensive phylogeny of the family Sparidae (Perciformes: Teleostei) inferred from mitochondrial gene analyses. *Genes Genet. Syst* **84**, 153-170.

Chistiakov DA, Hellemans B, Volckaert FAM (2006) Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture* **255**, 1-29.

Chopelet J, Waples RS, Mariana S (2009) Sex change and the genetic structure of marine populations. *Fish and Fisheries* **10**, 329-343.

Cowen RK, Paris CB, Srinivasan A (2006) Scaling of Connectivity in Marine Populations. *Science* **311**, 522-527.

De La Herran R, Rejon CR, Rejon MR, Garrido-Ramos MA (2001) The molecular phylogeny of the Sparidae (Pisces, Perciformes) based on two satellite DNA families. *Heredity* **87**, 691-697.

Devine JA, Baker KD, Haedrich RL (2006) Deep-sea fishes qualify as endangered. *Nature* **439**, 29.

Di Rienzo A, Peterson AC, Garza JC*, et al.* (1994) Mutational processes of simple-sequence repeat loci in human populations. *PNAS* **91**, 3166-3170.

Ellegren H (2004) Microsatellites: Simple sequences with complex evolution. *Nature reviews: Genetics* **5**, 435-445.

Fenberg PB, Roy K (2008) Ecological and evolutionary consequences of size-selective harvesting: how much do we know? *Molecular Ecology* **17**, 209-220.

Franch R, Louro B, Tsalavouta M*, et al.* (2006) A genetic linkage map of the hermaphrodite teleost fish *Sparus aurata* L. *Genetics* **174**, 851-861.

Freeland JR (2005) *Molecular Ecology*, John Wiley & Sons, Ltd.

Garcia SM, Rosenberg AA (2010) Food security and marine capture fisheries: characteristics, trends, drivers and future perspectives. *Philosophical Transactions of the Royal Society B* **365**, 2869-2880

González-Wangüemert M, Cánovas F, Pérez-Ruzafa A, Marcos C, Alexandrino P (2010) Connectivity patterns inferred from the genetic structure of white seabream (*Diplodus sargus* L.). *Journal of Experimental Marine Biology and Ecology* **383**, 23-31.

González-Wangüemert M, Pérez-Ruzafa A, Cánovas F, Garcia-Charton JA, Marcos C (2007) Temporal genetic variation in populations of Diplodus sargus from the SW Mediterranean Sea. *Marine Ecology Progress Series* **334**, 237-244.

Gonzalez EG, Beerli P, Zardoya R (2008) Genetic structuring and migration patterns of Atlantic bigeye tuna, *Thunnus obesus* (Lowe, 1839). *BMC Evolutionary Biology* **252**, 1-14.

Grant WS, Bowen BW (1998) Shallow Population Histories in Deep Evolutionary Lineages of Marine Fishes: Insights from Sardines and Anchovies and Lessons for Conservation. *The American Genetics Association* **89**, 415-426.

Gratwicke B, Speight MR (2005) The relationship between fish species richness, abundance and habitat complexity in a range of shallow tropical marine habitats. *Journal of Fish Biology* **66**, 650-667.

Griffiths MH (2000) Long-term trends in catch and effort of commercial linefish off South Africa's cape province: Snapshots of the 20th century. *South African Journal of Marine Science* **22**, 81-110.

Griffiths MH, Attwood C, Thomson R (1999) Appendix 1. New management protocol for South African Linefishery. 1-15.

Hansen MM, Kenchington E, Nielsen EE (2001) Assigning individual fish to populations using microsatellite DNA markers. *Fish and Fisheries* **2**, 93-112.

Hatanaka A, Yamada S-I (2006) Isolation and application of microsatellite DNA markers for pedigree tracing of seedlings of Red sea bream (*Pagrus major*). *Journal of the World Aquaculture Society* **37**, 139-143.

Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries* **9**, 333-362.

Hauser L, Ward RD (1998) Population identification in pelagic fish: the limits of molecular markers. *Advances in Molecular Ecology*, IOS Press.

Heemstra P, Heemstra E (2004) *Coastal Fishes of Southern Africa.* South African Institute for Aquatic biodiversity and National Inquiry Service Centre.

Hey J, Machado C (2003) The study of structured populations-new hope for a difficult and divided science. *Nature Reviews: Genetics* **4**, 535-543.

Hudson ME (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Resources* **8**, 1-15.

Hutchings L, van der Lingen CD, Shannon LJ*, et al.* (2009) The Benguela Current: An ecosystem of four components. *Progress in Oceanography* **83**, 15-32.

Hutchinson WF, Van Oosterhout C, Rogers SI, Carvalho GR (2003) Temporal analysis of archived samples indicates marked genetic changes in declining North Sea cod (*Gadus morhua*). *Proceedings of the Royal Society London B* **270**, 2125-2132.

Jean CT, Lee SC, Liu CW, Tseng MC (2006) Isolation and characterization of eight microsatellite loci from Picnic seabream (*Acanthopagrus berda*). *Molecular Ecology Notes* **6**, 1269-1271.

Kelly RP, Eernisse DJ (2007) Southern hospitality: a latitudinal gradient in gene flow in the marine environment. *Evolution* **61**, 700-707.

Kuparinen A, Merila J (2007) Detecting and managing fisheries induced evolution. *Trends in Ecology and Evolution* **22**, 652-659.

Leis JM, Trnski T, Beckley LE (2002) Larval development of *Pagellus natalensis* and what larval morphology indicates about relationships in the perciform fish family Sparidae (Teleostei). *Marine Freshwater Res* **53**, 367-376.

Li C, Riethoven J-JM, Ma L (2010) Exon-primed Intron-crossing (EPIC) markers for non-model teleost fish. *BMC Evolutionary Biology* **10,** 90.

Liu YG, Liu LX, Wu ZX*, et al.* (2007) Isolation and characterization of polymorphic microsatellite loci in black sea bream (*Acanthopagrus schlegeli*) by cross-species amplification with six species of the Sparidae family. *Aquatic Living Resources* **20**, 257-262.

Liu ZJ, Cordes JF (2004) DNA markers technologies and their applications in aquaculture genetics. *Aquaculture* **238**, 1-37.

Longhurst A (2007) Doubt and certainty in fishery science: Are we really headed for a global collapse of stocks? *Fisheries Research* **86**, 1-5.

Lutjeharms JRE, Monteiro PMS, Tyson PD, Obura D (2001) The oceans around southern Africa and regional effects of global change. *South African Journal of Science* **97**, 119-130.

Morin PA (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* **19**, 208-216.

Narum SR, Banks M, Beacham TD*, et al.* (2008) Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology* **17**, 3464-3477.

Navarro A, Badilla R, Zamorano MJ*, et al.* (2008) Development of two new microsatellite multiplex PCRs for three sparid species: Gilthead seabream (*Sparus auratus* L.), red porgy (*Pagrus pagrus* L.) and redbanded seabream (*P.auriga*, Valenciennes, 1843) and their application to paternity studies. *Aquaculture* **285**, 30-37.

Nicholson SE (2010) A low-level jet along the Benguela coast, an intergral part of the Benguela current ecosystem. *Climatic Change* **99**, 613-624.

Nielsen EE, Wright PJ, Hemmer-Hansen J*, et al.* (2009) Microgeographical population structure of cod *Gadus morhua* in the North Sea and west of Scotland: the role of sampling loci and individuals. *Marine Ecology Progress Series* **376**, 213-225.

O'Connell M, Wright JM (1997) Microsatellite DNA in fishes. *Reviews in Fish Biology and Fisheries* **7**, 331-363.

O'Connor MI, Bruno JF, Gaines SD*, et al.* (2007) Temperature control of larval dispersal and the implications for marine ecology, evolution, and conservation. *PNAS* **104**, 1266-1271.

Orrell TM, Carpenter KE (2004) A phylogeny of the fish family Sparidae (porgies) inferred from mitochondrial sequence data. *Molecular Phylogenies and Evolution* **32**, 425-434.

Patarnello T, Volckaert FAMJ, Castilho R (2007) Pillars of Hercules: is the Atlantic-Mediteranean transition a phylogeographical break? *Molecular Ecology* **16**, 4426-4444.

Pawson MG, Jennings S (1996) A critique of methods for stock identification in marine capture fisheries. *Fisheries Research* **25**, 203-217.

Perez L, Infante C, Ponce M*, et al.* (2008) Characterization of eight microsatellite markers in the white sea bream, *Diplodus sargus* (Teleostei, Sparidae). *Molecular Ecology Resources* **8**, 1291-1293.

Piñera JA, Bernardo D, Blanco G, Vazquez E, Sanchez JA (2006) Isolation and characterization of polymorphic microsatellite markers in *Pagellus bogaraveo*, and cross-species amplification in *Sparus aurata* and *Dicentrarchus labrax*. *Molecular Ecology Notes* **6**, 33-35.

Piñera JA, Bernardo D, Blanco G, Vazquez E, Sanchez JA (2007) Usefulness of microsatellite markers developed from Pagellus bogaraveo to genetically study five different species of Sparidae. *Marine Ecology* **28**, 184-187.

Planes S, Lenfant P (2002) Temporal change in the genetic structure between and within cohorts of a marine fish, *Diplodus sargus*, induced by a large variance in individual reproductive success. *Molecular Ecology* **11**, 1515-1524.

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nature reviews Genetics* **6**, 1-13.

Ponce M, Infante C, Crespo A*, et al.* (2006) Characterization of microsatellite loci for the redbanded seabream, *Pagrus auriga* (Teloestei, Sparidae). *Molecular Ecology Notes* **6**, 527-529.

Primmer CR, Painter JN, Koskinen MT, Palo JU, Merila J (2005) Factors affecting avian cross-species microsatellite amplification. *Journal of Avian Biology* **36**, 348-360.

Pulfrich A, Griffiths CL (1988a) The Fishery for Hottentot Pachymetopon blochii Val. in the South-Western Cape. *South African Journal of Marine Science* **7**, 227-241.

Pulfrich A, Griffiths CL (1988b) Growth, sexual maturity and reproduction in the hottentot *Pachymetopon blochii* (Val.). *South African Journal of Marine Science* **7**, 25-36.

Punt AE, Pulfrich A, Butterworth DS, Penney AJ (1996) The effect of hook size on the size-specific selectivity of Hottentot *Pachymetopon blochii* (Val.) and on yield per recruit. *South African Journal of Marine Science* **17**, 155-172.

Rico C, Rico I, Hewitt G (1996) 470 Million years of Conservation of Microsatellite Loci among Fish Species. *Proceedings of the Royal Society London: Biological Sciences* **263**, 549-557.

Roberts DG, Ayre DJ (2010) Panmictic population structure in the migratory marine sparid *Acanthopagrus australis* despite its close association with estuaries. *Marine Ecology Progress Series* **412**, 223-230.

Rochet M-J (2000) A comparative approach to life-history strategies and tactics among four orders of teleost fish. *ICES Journal of Marine Science* **57**, 228-239.

Roques S, Galarza JA, Macpherson E*, et al.* (2007a) Isolation of eight microsatellite loci from the saddled bream *Oblada melanura* and cross-species amplification in two sea bream species of the genus *Diplodus*. *Conservation Genetics* **8**, 1255-1257.

Roques S, Galarza JA, Macpherson E, Turner GF, Rico C (2007b) Isolation and characterization of nine polymorphic microsatellite markers in the two-banded sea bream (*Diplodus vulgaris*) and cross-species amplification in the white sea bream (*Diplodus sargus*) and the saddled bream (*Oblada melanura*). *Molecular Ecology Notes* **7**, 661-663.

Ryman N, Utter F, Laikre L (1995) Protection of intraspecific biodiversity of exploited fishes. *Reviews in Fish Biology and Fisheries* **5**, 417-446.

Sakko AL (1998) The influence of the Benguela upwelling system on Namibia's marine biodiversity. *Biodiversity and Conservation* **7**, 419-433.

Santana QC, Coetzee MPA, Steenkamp ET*, et al.* (2009) Microsatellite discovery by deep sequencing enriched genomic libraries. *BioTechniques* **46**, 217-223.

Sauer WHH, Hetch T, Britz PJ, Mather D (2003) An economic and sectoral study of the South African fishing industry. *Fishery Profiles* **2**, 1-312.

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**, 615-629

Shanks AL, Eckert G (2005) Population persistence of California current fishes and benthic crustaceans: a marine drift paradox. *Ecological Monographs* **75**, 505-524.

Sheaves M (2006) Is the timing of spawning in sparid fishes a response to sea temperature regimes? *Coral Reefs* **25**, 655-669.

Soulsbury CD, Iossa G, Edwards KJ (2009) The influence of evolutionary distance between cross-species microsatellites and primer base-pair composition on allelic dropout rates. *Conservation Genetics* **10**, 792-802.

Teske PR, Forget FRG, Cowley PD, von der Heyden S, Beheregaray LB (2010) Connectivity between marine reserves and exploited areas in the philopatric reef fish Chrysoblephus laticeps (Teleostei: Sparidae). *Marine Biology* **157**, 2029-2042.

Tóth G, Gáspári Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research* **10**, 967-981.

Turpie JK, Beckley LE, Katua SM (2000) Biogeography and the selection of priority areas for conservation of South African coastal fishes. *Biological Conservation* **92**, 59-72.

Turpie JK, Heydenrych BJ, Lamberth SJ (2003) Economic value of terrestrial and marine biodiversity in the Cape floristic region: implications for defining effective and socially optimal conservation strategies. *Biological Conservation* **112**, 233-251.

Van der Elst R (1988) *A guide to the common sea fishes of Southern Africa*, 2nd edition. C. Struik Publishers.

Wagner AP, Creel S, Kalinowski ST (2006) Estimating relatedness and relationships using microsatellite loci with null alleles. *Heredity* **97**, 336-345.

Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity* **89**, 438-450.

Waples RS (2010) Spatial-temporal stratifications in natural populations and how they affect understanding and estimation of effective population size. *Molecular Ecology Resources* **10**, 785-796.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their connectivity. *Molecular Ecology* **15**, 1419-1439.

Ward RD (2000) Genetics in fisheries management. *Hydrobiologia* **420**, 191-201.

Weersing K, Toonen RJ (2009) Population genetics, larval dispersal, and connectivity in marine systems. *Marine Ecology Progress Series* **393**, 1-12.

Worm B, Barbier EB, Beaumont N, *et al.* (2006) Impacts of Biodiversity Loss on Ocean Ecosystem Services. *Science* **314**, 787-790.

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology* **11**, 1-16.

Zhan A, Hu J, Hu X*, et al.* (2009) Fine-scale population genetic structure of zhikong scallop (Chlamys farreri): Do local marine currents drive geographical differentiation? *Marine Biotechnology* **11**, 223-235.

Chapter 2

Isolation and characterization of microsatellites in two South African endemic sparids *Pachymetopon blochii* (Valenciennes 1830) and *Lithognathus lithognathus* (Curvier 1829) and cross-species amplification in nine other economically important sparids

Abstract

The hottentot seabream (*Pachymetopon blochii*) and white steenbras (*Lithognathus lithognathus*) are endemic sparids occurring along the coast of southern Africa and are important line fish for recreational and subsistence fishermen. Population genetic studies can make a valuable contribution to improve understanding of the population biology of these species. To enable such fine-scale, nine polymorphic microsatellite loci were isolated in each of these species using the Fast Isolation by AFLP of Sequences COntaining repeats (FIASCO) protocol and 454 Life Sciences/Roche GS-FLX sequencing. In the hottentot seabream the number of alleles (N_A) and observed heterozygosity (H_O) ranged from 11 - 30 and 0.408 - 0.940, respectively. For the white steenbras N_A ranged from 5 - 15 and H_O from 0.263 - 0.940. After subsequent cross-species testing a total of 13 markers (nine developed and four from other sparids) were identified in the white steenbras and 16 (nine developed and seven from cross-species testing) were found to be polymorphic in the hottentot seabream. Nineteen markers (fifteen from the present study and four from the literature) were further tested for cross-species amplification and polymorphism in nine other economically important sparids found around the coast of South Africa. Polymorphic markers were identified in all species and ranged from five in *Chrysoblephus puniceus* to eleven in *Diplodus sargus capensis*. In the Sparidae, amplification success and polymorphism do not appear to be negatively correlated with genetic distance between the source and target species. Future studies can employ these markers as a genetic tool for identifying the population genetic structure of these species occurring along the coast of South Africa.

Keywords: 454 pyrosequencing, cross-species amplification, *Lithognathus lithognathus*, microsatellites, *Pachymetopon blochii*, Sparidae

2.1 Introduction

Microsatellites are tandem repeat-containing loci, which are co-dominant, generally selectively neutral, often highly variable and are found throughout eukaryotic genomes (Ellegren 2004; Selkoe & Toonen 2006). These markers have been extensively used to identify the underlying population structure of marine fish species as they enable the inference of gene flow across multiple scales (Selkoe & Toonen 2006). Microsatellite loci are thought to mutate either through slippage of the DNA polymerase or unequal recombination events leading to high levels of polymorphism (Di Rienzo *et al.* 1994; Zane *et al.* 2002).

Microsatellite markers in non-model organisms are generally identified *de novo* through time consuming and relatively expensive methods. Zane *et al.* (2002) reviewed numerous development processes, including creating selective hybridization partial genomic libraries enriched for repeat fragments (Fast Isolation by AFLP of Sequences COntaining repeats; FIASCO), sequencing of expressed sequence tags (ESTs) and primer extension (ISSR-PCR; Zane *et al.* 2002). With the advent of next-generation sequencing (high throughput methods) the time of cloning and sequencing selective hybridization partial genomic libraries may be saved. Recent studies have used 454 LifeSciences/Roche GS-FLX sequencing to directly sequence the repeat-enriched DNA isolated through the above mentioned protocols (Allentoft *et al.* 2009; Santana *et al.* 2009; Castoe *et al.* 2010). This method has proven to be cost effective per sequence and time saving (Santana *et al.* 2009; Castoe *et al.* 2010).

An alternative method to isolating microsatellites *de novo* is to test microsatellites isolated from a source species in closely related species (i.e. cross-species amplification) due to conservation of the flanking regions and thus the priming sites (Zane *et al.* 2002). Microsatellites identified in marine fish have shown extensive flanking region conservation in some of the markers, with one locus being conserved over 470 million years in a number of different fish families (Rico *et al.* 1996). Other studies, however, have identified clear negative correlations between genetic distance, amplification success and polymorphism in several taxa of vertebrates including amphibians, birds, cetaceans and marine fish (Primmer *et al.* 2005; Barbara *et al.* 2007; Carreras-Carbonell *et al.* 2008). Carreras-Carbonell and co-workers (2008) determined a correlation from previous microsatellite cross-species studies of marine fish and the percent divergence (i.e. genetic distance) between taxa based on the 16S or 12S mitochondrial rDNA genes. Their study observed a general negative correlation between the genetic divergence of species and the success of microsatellite amplification and polymorphism. This trend was observed across all the families except sparids. However, only a small data set for sparids was included in their study and this will be expanded on in

this study. This correlation equation may be used to predict the degree to which the microsatellites from a source species will amplify and be polymorphic in a species of interest (Carreras-Carbonell *et al.* 2008).

The advantage of using cross-species amplification is that it saves the time and cost of *de novo* development (Carreras-Carbonell *et al.* 2008). Mutations in the primer binding sites of non-source species may lead to null alleles or non-specific amplification (Liu *et al.* 2007; Soulsbury *et al.* 2009). Thus before using a set of markers identified through cross-species amplification, each marker should be tested for deviations from Hardy-Weinberg equilibrium (HWE) and null alleles (Liu *et al.* 2007). If some artefacts are detected, primers may be re-designed or the marker should be discarded. Many fish studies have reported markers that amplify in other closely related species, particularly in the Sparidae where many cross-species amplification studies have been conducted (Brown *et al.* 2005; Chen *et al.* 2005; Piñera *et al.* 2006; Ponce *et al.* 2006; Castro *et al.* 2007; Liu *et al.* 2007; Piñera *et al.* 2007; Roques *et al.* 2007a; Roques *et al.* 2007b; Pérez *et al.* 2008).

Along the South African coast the Sparidae is one of the most economically important fish families. Twenty-four of the 110 sparids recorded along the South African coast are endemic to the region (Van der Elst 1988; Heemstra & Heemstra 2004). Two endemic species of recreational and subsistence importance are the hottentot seabream (*Pachymetopon blochii*) and the white steenbras (*Lithognathus lithognathus*). *Pachymetopon blochii* is distributed along the south and west coast of southern Africa, from Port Alfred to Angola, where it is mainly found in kelp beds and rocky outcrops (Heemstra & Heemstra 2004). This species is considered vulnerable as the adults are relatively sedentary (so they are more easily targeted) and sexually mature at approximately five years (Heemstra & Heemstra 2004). The white steenbras has a distribution range from the Tugela river on the east coast of South Africa to Angola (Heemstra & Heemstra 2004). In recent years the catches of white steenbras have significantly declined due to degradation of estuaries (juvenile habitat) and increasing fishing pressure (Bennett 1993; Heemstra & Heemstra 2004).

The Sparidae are of economic importance for South African fisheries and require comprehensive population genetic studies with competent microsatellite markers. The aim of this study was to develop microsatellite markers (*de novo*) for the hottentot seabream and white steenbras through the FIASCO protocol and 454 sequencing. A cross-species study was then conducted across 11 sparids using the *de novo* microsatellites and previously developed markers in sparids as a panel of loci. The relationship between microsatellite reliability and genetic distance was evaluated with this data.

2.2 Materials and Methods

2.2.1 Sampling and DNA extraction

For characterization of the microsatellite markers 50 samples of *P. blochii: 25* from Sewejaarskop se mond (17°73'07" -31°13'03") and 25 samples from Danger Point (19°30'30" -34°63') were extracted with the DNeasy Blood and tissue extraction kit (Qiagen). Fifty *L. lithognathus* samples from Knysna (14), East Kleinemonde (17) and Sundays Estuary (19) were extracted with the Wizard Genomic DNA Purification Kit (Promega). Sixteen samples each of *Acanthopagrus berda*, *Cheimerius nufar, Chrysoblephus laticeps*, *Chrysoblephus puniceus*, *Diplodus cervinus hottentotus, Diplodus sargus capensis*, *Polysteganus undulosus*, *Rhabdosargus globiceps*, and *Rhabdosargus holubi* were extracted with the DNeasy Blood and tissue extraction kit (Qiagen) for cross-species testing.

2.2.2 Development of microsatellites for hottentot seabream and white steenbras

A repeat-enriched genomic fraction was generated for both the white steenbras and hottentot seabream using the FIASCO protocol developed by Zane *et al*. (2002). A simultaneous digestion/ligation reaction was performed with 250 ng of genomic DNA using *Mse*I restriction enzyme (New England Biolabs) and *Mse*I AFLP adaptors (5'-TACTCAGGACTCAT-3'/5'-GACGATGAGGTCCTGAG-3'; Zane *et al.* 2002). A polymerase chain reaction (PCR) was prepared with primers specific to the *Mse*I adaptors (*Mse*I-*N*; 5'-GATGAGTCCTGAGTAAN-3') as described by Zane *et al*. (2002). A biotinylated probe combination containing tetranucleotide repeats of $CAAA_6$, $GAAA_6$ and $GATA_6$ was hybridized with the amplified DNA fragments to enrich for repeat motifs. The repeat-containing DNA was then captured with streptavidin-coated beads (Roche).

Five non-stringent washes were performed by washing the DNA-probe-bead complex with TEN_{1000} (10mM Tris-HCl, 1 mM EDTA, 1 M NaCl, pH7.5) and then the complex was separated from the buffer by applying a magnetic field. This was followed by five stringent washes (0.2 x SSC, 0.1 % SDS), and heat (TE and 95 °C) and chemical denaturation (0.15 M NaOH) to separate the DNA fragments containing repeats from the streptavidin-coated beads and probes. The enriched DNA was precipitated and PCRs were carried out as described by Zane *et al*. (2002) to amplify approximately 2 µg of DNA. This enriched DNA was purified using the High Pure PCR Product Purification Kit (Roche).

The repeat-enriched DNA was sequenced using 454 LifeSciences/Roche GS-FLX pyrosequencing (Inqaba Biotec, Pretoria, South Africa). Each of these species had approximately one fifth of a lane allocated for sequencing. The 454 sequencing (done by Inqaba Biotec) included the annealing of short A and B adaptors to each sequence and the denaturation of dsDNA; these adaptors are required for the subsequent steps and only single stranded sequences with both adaptors annealed comprise the partial genomic library. These sequences are then mixed with Sepharose beads which either contain an A or B adaptor. The DNA fragments hybridize to the beads and then an emulsion PCR is done following manufacturer's specifications. The beads containing the amplified sequences are transferred to a PicoTiterPlate™ for sequencing by the Genome Sequencer FLX™ where they are sequenced (Droege & Hill 2008).

The program MSATCOMMANDER (Faircloth 2008) was used to identify sequences containing repeat motifs from the 454 sequences of both species. Potential loci with tri- and tetranucleotide repeats were detected and selected for primer development with PRIMER 3 version 1.1.1 (Rozen & Skaletsky 2000).

A panel of 16 individuals were used to optimise PCR amplification and identify polymorphic loci in each species. PCR reactions were prepared in 10 µl volumes with approximately 50 ng of genomic DNA, 1x PCR buffer (100 mM Tris-HCl, 500 mM KCl, pH 8.4), 0.4 U Supertherm *Taq* polymerase (Southern Cross Biotechnologies), 1.5 mM $MgCl_2$, 0.1 pmol of the forward and reverse primer (Inqaba Biotec) and 0.2 mM dNTPs (Promega). The PCR conditions were: 94 ^{o}C for 5 min followed with 94 ^{o}C for 30 s, 56 – 61 ^{o}C for 45 s and 72 ^{o}C for 45 s for 35 cycles and a final elongation step of 72 ^{o}C for 20 min followed. The PCR products were electrophoresed on 2 % agarose gels and visualized with GelRed™ Acid stain (Biotium). For optimised loci the PCR reaction was repeated with 0.02 pmol fluorescently labelled ChromaTide Rhodamine Green™-5-dUTPs (Invitrogen); these are labelled dNTPs which are incorporate during the PCR reaction and then allow the PCR fragments to be observed on an automated sequencer. This was analysed on an ABI 3100 automated sequencer (Applied Biosystems, Foster City, CA) with the GeneScan Liz™ 500 Size Standard (Applied Biosystems).

The forward primers of loci that were found to be polymorphic were fluorescently labelled with the G5 dye set including 6-FAM, VIC, NED and PET (Applied Biosystems). These loci were genotyped in 50 individuals of each species (hottentot seabream and white steenbras) using the Quantitect Multiplex PCR kit (Qiagen) following the criteria recommended by the manufacturer. Approximately 20 ng of genomic DNA, half the final volume Quantitect

multiplex master mix, 0.2 pmol of each labelled forward and reverse primer for each primer set were included in the multiplex PCR reaction (10 µl). The PCR amplification protocol entailed 95 °C for 15 minutes followed by 45 cycles of 94 °C for 60 s and 60 °C for 90 s. The multiplex combinations are indicated in Table 2.1A and 2.1B. Forty hottentot and 20 white steenbras samples were amplified twice for all microsatellite loci to determine genotyping error.

2.2.3 Sequencing of homozygotes

Homozygotes for a selection of alleles from both the hottentot seabream and white steenbras were amplified using the protocol mentioned above (using unlabeled primers) and then sequenced to confirm repeat-motif variation and allele scoring. The DNA amplification product was precipitated using 3 volumes absolute ethanol, 0.4 volumes Sabax® water (Adcock-Ingram) and 0.08 volumes 3 M sodium acetate (NaAc) The precipitated product was eluted in 15 µl Sabax® water and confirmed on a 2 % agarose gel visualized with GelRed. The precipitated DNA product was sequenced in the forward orientation only with the forward primers indicated in the amplification step (Table 2.1A and 2.1B). The ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit version 3.1 (Applied Biosystems) was used to perform the cycle sequencing reaction on an ABI GeneAMP® 9700 PCR system. The reactions contained 20 - 80 ng of purified template DNA, 2 µl of the BigDye reaction mix and 0.5x BigDye sequencing buffer. These were cycle sequenced using the manufacturer's recommended conditions and precipitated as described above. Sequences were analysed on an ABI 3130 automated sequencer (Applied Biosystems). The chromatograms were visualized with BioEdit version 7.0.2 (Hall 1999).

2.2.4 Cross-species amplification

Markers identified to be polymorphic in the hottentot and white steenbras were used for evaluating the success of cross-species amplification. The species that were included are *A. berda*, C. *nufar, C. laticeps*, *C. puniceus, D. c. hottentotus*, *D. s. capensis*, *P. undulosus, R. globiceps* and *R. holubi*. These species were included as they are of economic and current research importance and they represent a spread across the sparid phylogeny shown in Fig. 2.1 (Chiba *et al.* 2009). Newly developed markers that showed inconsistent amplification or significant deviation from HWE were not included further in this study. Markers included *P. blochii* were PBt001, PBt003, PBt007, PBt011, PBt012, PBt014, PBt018 and *L. lithognathus* LLt002, LLtr004, LLt005, LLt006, LLt007, LLt011, LLt014, LLt020.

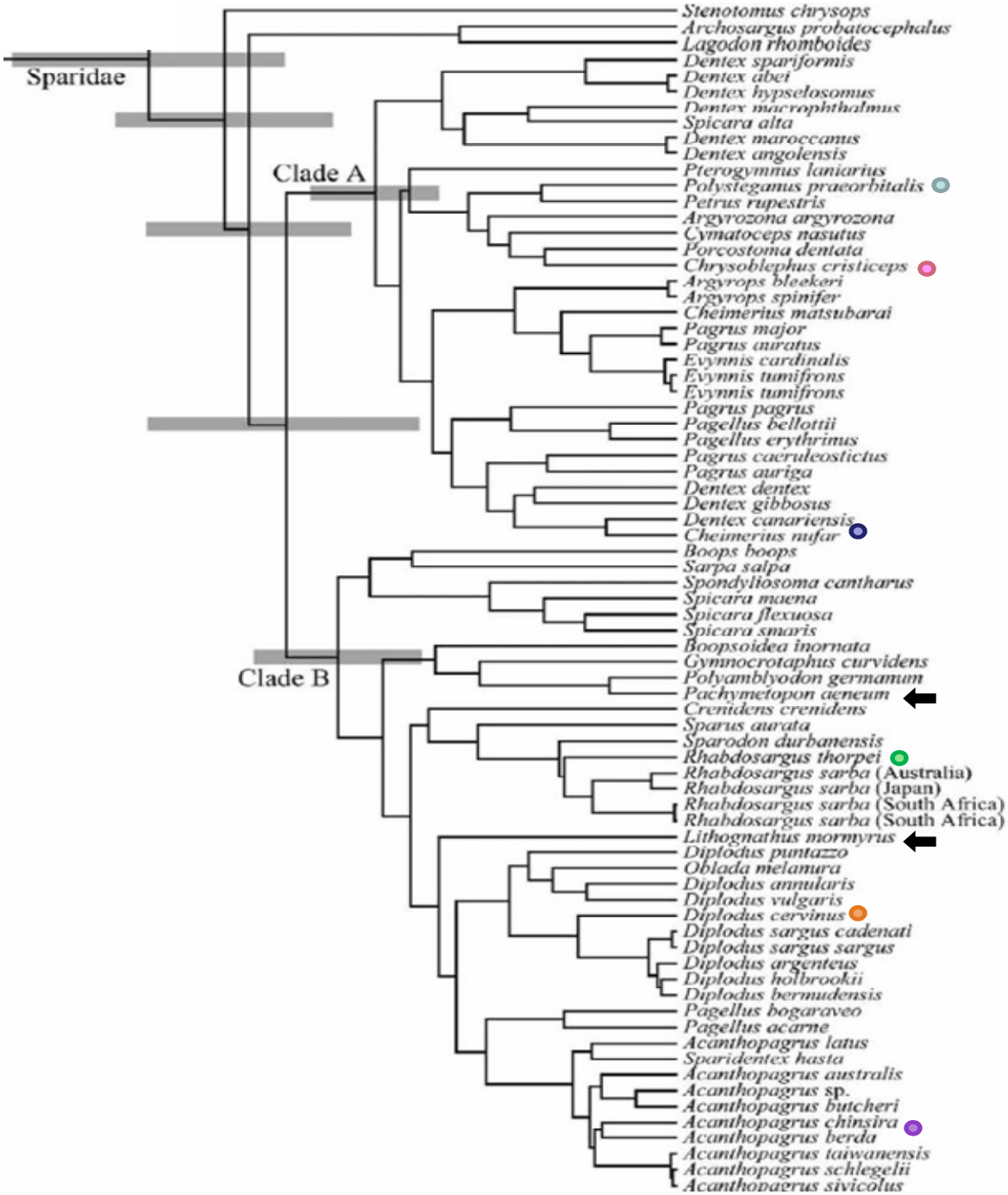


Fig. 2.1 Sparid phylogeny based on the mitochondrial DNA cytochrome *b* gene. The coloured dots indicate the genera represented in the present cross-species amplification tests, their position in the phylogeny and the clade they belong to. The two focal genera are indicated by black arrows. This figure was modified from Chiba *et al.* (2009). The two main species included in this study were not included in the phylogeny.

In addition four tetranucleotide microsatellites identified through previous cross-species amplification tests were included for further analysis; these were found to be polymorphic in either or both focal species. These markers included PA535 (*Pagrus auriga*) and SA043 (*Sparus.* aurata) where primers were designed from the original sequences of unpublished repeats (PA535, AB242995 F:GTCGAGGCTGTAAACAGGAC, R:CGTGGCTTTGGTTATTTCTTGC and SA043 , DQ851346 F:TGCATTATATCACACAACACACG, R:TGGACCACCAGACCAGTTAG). Primers Pb-OVI-D106 (herein referred to as PB106; Piñera *et al.* 2006) and Clat11 (referred to as CL011; Teske *et al.* 2009) were identified from primer notes. Also, cross-species testing of the additional markers was conducted in the hottentot and white steenbras to increase the total number of markers for population genetic studies. These markers were fluorescently labelled: PA535 (6-FAM), SA043 (NED), PB106 (VIC) and CL011 (VIC). Combinations of four markers with differing fluorescent colours were amplified in four individuals with the Qiagen Multiplex Kit as recommended by the manufacturers. Markers that amplified consistently were subsequently genotyped in 16 individuals of each species to determine levels of polymorphism.

2.2.5 Data analysis of microsatellites

Genotypes were scored using GENEMARKER version 1.5 software (SoftGenetics, State College, Pennsylvania, USA). Observed heterozygosity (H_O), expected heterozygosity (H_E) and F_{IS} were calculated with GENETIX version 4.05 (Belkhir *et al.* 1996-2004). Deviations from Hardy-Weinberg equilibrium (HWE) were calculated with GENEPOP version 4.0 (Rousset 2008) for the developed markers and GENETIX version 4.05 for the cross-species section. Linkage disequilibrium was determined using ARLEQUIN version 3.11, with 100 000 Markov Chain steps (Excoffier *et al.* 2005). The incidence of null alleles, scoring errors due to stuttering and evidence of large allele drop-out were determined using MICRO-CHECKER version 2.2.3 (van Oosterhout *et al.* 2004). A Mann-Whitney test was done between the number of alleles of *P. blochii* and *L. lithognathus* from the *de novo* developed loci to compare levels of polymorphism between the two species. This analysis was conducted as a one-tailed test with the program STATEXT version 1.4.2b (www.statext.com)

2.2.6 Evaluation of cross-species amplification and polymorphism

Sequences for the mitochondrial 16S rDNA gene were downloaded from GenBank for the species included in this study. Sequences were not available for all species, thus species in the same genus were included for those species to provide an estimate of sequence divergence. The species names and GenBank accession numbers are included in Appendix 2A. These sequences were aligned in Clustal X (Thompson *et al.* 1997) and the genetic divergence was estimated as the percentage of haplotype sequence differences, which was calculated in PAUP version 4.0b10 (Swofford 2001). The genetic divergence values were then used in the equations derived by Carreras-Carbonell *et al.* (2008) for 16S to estimate the percentage microsatellite amplification ($y=105.6-58\text{Log}_{10}(x)$, where x is the genetic distance) and polymorphism ($y=116.8-84.8\text{Log}_{10}(x)$), according to genetic divergence in marine fish; The equation is derived from a negative correlation identified across 10 marine fish families. The predictions were compared to the observed data to identify whether the included members of the sparid family showed a negative correlation between genetic divergence and amplification success and polymorphism, which would agree in general with other marine fish. If such a negative correlation exists, this calculation may be used for predicting which species will be the best source species for cross-species amplification in a species of interest (Carreras-Carbonell *et al.* 2008). The significance of the correlation was tested by converting the actual AL and PL values of each species to a logarithmic scale and then testing a simple linear regression using the program STATEXT version 1.4.2b (www.statext.com).

2.3 Results

2.3.1 Sequencing results from 454 sequencing

The number of sequences obtained from the hottentot seabream library was 11076, of which 4531 (41 %) sequences contained di-, tri- or tetranucleotide repeats with more than four repeat units (here a microsatellite was considered as having > four tandem repeats). From the white steenbras there was a total of 7148 sequences and of these 2839 (39 %) contained either di-, tri- or tetranucleotide repeats. Although only tetranucleotide sequences were probed for, the probes also captured di-, tri- and pentanucleotide repeats. The fragments ranged from 40 - 600 base pairs (bp) in length with the majority of sequences being between 120 and 400 bp for both the hottentot and white steenbras (Fig. 2.2A). A comparison between the number of sequences containing six or more tandem repeats of each type of repeat for the hottentot and white steenbras probing is provided in Fig 2.2B. This repeat

length was set as a minimum for primer design as polymorphism is more likely for longer repeat lengths (Buschiazzo & Gemmell 2006). A one-fold coverage of each sequence was obtained.

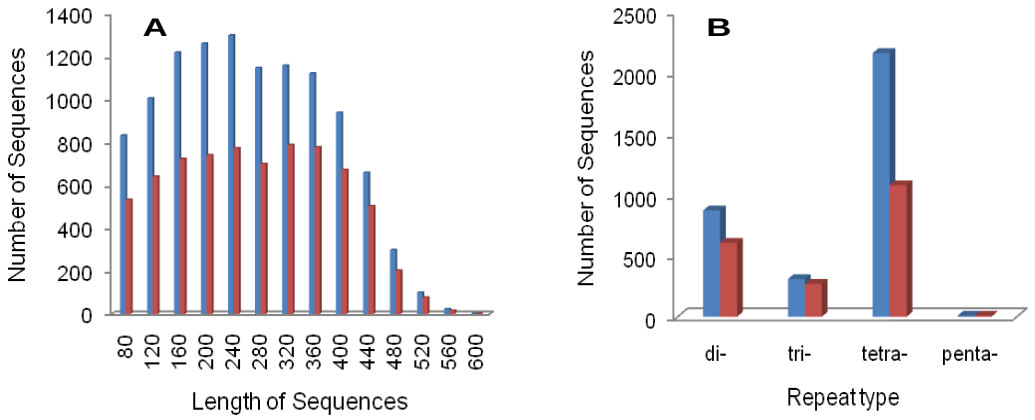


Fig. 2.2 A) Distributions of size lengths of the fragments obtained from 454 lifeSciences/Roche GS-FLX sequencing for *P. blochii* (blue) and *L. lithognathus* (red) after repeat enrichment. **B**) Comparison of the number of sequences for each species containing six repeats or more of each repeat type.

2.3.2 Isolation and characterisation of microsatellites

The first 20 sequences containing tetranucleotide repeats with sufficient flanking region for primer design were selected for *P. blochii* and 12 potential loci amplified the expected fragment size at 61 ^{o}C. Nine of these loci were identified as polymorphic through Genescan analysis using fluorescently labelled dNTPs (Table 2.1A; Appendix 2B). Sequences for polymorphic loci will be submitted to GenBank. The re-genotyped loci in *P. blochii* showed no evidence of genotyping errors. No linkage disequilibrium was identified between these markers, suggesting that they are all physically unlinked and therefore all informative. The microsatellite loci PBt018 and PBt006 significantly deviated from HWE. The program MICRO-CHECKER detected the presence of null alleles at two loci, with frequencies estimated as 6.54 % and 25.73 % for PBt018 and PBt006, respectively. The program suggested evidence of scoring error due to stuttering or large allele drop-out according to MICRO-CHECKER. The observed heterozygosity (H_O) ranged from 0.408 for marker PBt006 to 0.940 for PBt011 (Table 2.1A).

Forty primer sequences were designed for tri- and tetranucleotide repeat sequences in *L. lithognathus;* this included all tri- and tetranucleotide sequences that had sufficient flanking region. Nineteen primer pairs were optimised and showed the correct size fragment and nine loci were polymorphic (Appendix 2B). In this species all loci amplified consistently without non-specific amplification at 61 ^{o}C. The characterisation of these markers is summarised in Table 2.1B. There was no evidence of error from the re-genotyping and no linkage disequilibrium between the loci, which indicated that they were likely informative. Marker LLt024 did not amplify consistently and showed a significant homozygote excess. Markers that showed potential null alleles were LLt005 (11.43 %) and LLt024 (33.24 %) according to the Oosterhout estimate (van Oosterhout *et al.* 2004). MICROC-CHECKER indicated that LLt024 showed scoring errors which could be explained by the presence of stuttering. The H_O ranged from 0.263 for LLt024 to 0.940 for LLt020 (Table 2.1B). A significant difference was identified through the Mann-Whitney test between the number of alleles identified in the de novo developed microsatellite loci of the white steenbras and hottentot seabream (*P*-value = 0.002)

2.3.3 Sequencing of homozygotes

Sequencing of the forward strand was not done for PBt006 and LLtr024 due to inconsistent amplification and the high percentage of null alleles at these loci. Sequencing errors were identified in the 454 sequencing once a selection of the homozygotes were Sanger sequenced for the remaining microsatellite loci of *P. blochii* and *L. lithognathus*. These errors were found for multiple loci at A/T repeats; this included either a single base missing (PBt001, PBt013, PBt014, LLt007, LLt011 & LLtr004) or one extra base pair added (PBt018 & LLt011) in the 454 sequencing (Appendix 2B & 2C). From the Sanger sequencing of alleles, all markers showed no mutations in the flanking regions and the varying lengths of the alleles were due to variation in the number of repeats (Appendix 2C). The exception was marker PBt012, which showed a high level of variability in the flanking region and the allele length did not match the repeat unit length. This was specifically because of a polymorphic mononucleotide repeat and an insertion of eight base pairs in one allele (Appendix 2C).

Table 2.1A Isolated microsatellite loci from 50 *Pachymetopon blochii* samples with characterization of the primer sequence, repeat motif, allele size range, Observed (H_O) and Expected (H_E) heterozygosity and F_{IS} (Weir & Cockerham, 1984).

Locus	Dye	Primer sequence (5'-3')	Repeat motif	Allele range (bp)	N_A	H_O	H_E	F_{IS}
PBt001[3]	PET	F: AGG TTC CCC ACA GAA GGT C R: AGT ACC TGG GAA ACA GCC C	$(AGAT)_9$	209-265	13	0.900	0.848	-0.051
PBt003[3]	PET	F: TTG GGG AAA AGG AGA GGC G R: CCG ACT CGC TCT GTA TGT TG	$(GATA)_{17}$	143-203	15	0.840	0.913	0.090
PBt006[1]	6-FAM	F: CAT TTT GGC AGT CCG TCC G R: AAA TGT ACT GAC GTT CGT ATC C	$(AGAT)_{11}$	213-253	11	0.408	0.859	0.533***
PBt007[1]	6-FAM	F: CGC AGG ATT CAT ACG TAA AGA AAT R: TTC CAT TCC ATG CCC TGC C	$(TATC)_{13}$	130-178	13	0.880	0.888	0.020
PBt011[3]	NED	F: CAC AGA CGG GAA TGC ACA C R: AGG GTG TCA GGA CAG AAC G	$(TATC)_{17}$	183-271	20	0.940	0.914	-0.018
PBt012[2]	PET	F: CAC GCA CTG GTA TCG AAT CTG R: AGA GGT GGA GCT AGT CTG TC	$(TATC)_{12}$	144-292	30	0.920	0.944	0.031
PBt013[2]	VIC	F: GAG AAA CAA CAG TGT CAG GGG R: GCA AGA GTG ATG ACA CAC AGG	$(GATA)_3(GACA)_1(GATA)_{13}$	118-194	18	0.820	0.859	0.056
PBt014[3]	VIC	F: GTT GGG TGA TGG AGC CAA G R: CAG AGG TGA CTA GAA GCC CC	$(GATA)_{18}$	149-201	13	0.860	0.875	0.027
PBt018[1]	VIC	F: GTC AAG CCT GTC AAA CTG GG R: TGG AAA GAA ACA CGA GCC G	$(TATC)_{14}$	180-284	20	0.800	0.918	0.139*

P-value < 0.05*, *P*-value < 0.001***, [1,2,3] Indicates in which multiplex the markers were combined.

Table 2.1B Isolation of microsatellites from 50 white steenbras (*Lithognathus lithognathus*) samples, including primer sequence, repeat motif, allele size range, number of alleles (N_A), Observed (H_O), Expected heterozygosity (H_E) and F_{IS} (Weir and Cockerham, 1984).

Locus	Dye	Primer sequence (5'-3')	Repeat motif	Allele range (bp)	N_A	H_O	H_E	F_{IS}
LLt002[3]	VIC	F: GCTTTGACCGTCTTTTCCCC R: GTT TTG GGC TGT TGG TCG G	$(CTTT)_8$	242-286	12	0.900	0.877	-0.016
LLt005[1]	6-FAM	F: GTG AGA ATC CTC CAT CCA AGT C R: AAT GGG CCT TTC CAT GCT G	$(CAAA)_{11}$	219-235	5	0.520	0.678	0.243*
LLt006[1]	VIC	F: TCA TCT GAC CAC AGA CCA GC R:CAC GCC ATT TGT GAC AGT AAG	$(GATA)_{14}$	197-237	10	0.860	0.866	0.018
LLt007[2]	6-FAM	F: TGG ATA CCT GAA TGT CCC TCG R: GTC ATA CCC TGG GCT CCA C	$(TATC)_{13}$	181-237	15	0.860	0.843	-0.010
LLt011[1]	PET	F: AAA CCC ATC ACC CAG TCC C R: AAA ACA CTA CAG ATT AGA CGC TG	$(GATA)_{12}$	227-271	12	0.860	0.837	-0.016
LLt014[1]	PET	F: ATA CAA ATG CCA CTA GAC CAA C R: CTA TAG TAA GAT TGG TAG TCA GTG ATG	$(AGAT)_{10}$	140-172	9	0.820	0.840	0.035
LLt020[3]	6-FAM	F:CTG TGC CTT TAT CTG TAT CTA CTC R: TTG TCC ATC ACG ATG TTT C	$(CAAC)_{11}(CAAA)_{13}$	238-318	14	0.940	0.897	-0.038
LLt024[2]	NED	F: TGA CAG AAT GAA CGT CGG C R: GGT TCT GTT TCT GAG GTC TGG	$(AGAT)_{18}$	183-251	10	0.263	0.842	0.694***
LLtr004[2]	PET	F: CGC TCA TTG ATT GGA GCC C R: ACA CTG GAA CGA AAC TGG TTG	$(CAG)_{11}$	198-213	6	0.720	0.761	0.065

P-value < 0.05*, *P*-value < 0.001***. The numbers above the Locus name [1,2,3] indicate the multiplex combinations.

2.3.4 Cross-species amplification

Fifteen newly developed and the four additional microsatellite markers were tested in 11 sparid species: Seventeen of the loci are from clade A species and two from clade B (clade A and B; Fig. 2.1). The genera incorporated in the cross-species amplification are indicated in Fig. 2.1, illustrating the spread of species included in this study across the phylogeny. The characteristics and summary statistics, including N_A, H_O, H_E and allele size range and deviation from HWE, are summarised in Table 2.2 and Appendix 2D. The average number of polymorphic markers was 5.6 markers per species for clade B, and 4 polymorphic markers per species for clade A. Markers PA535, CL011, PBt003 and LLtr004 were polymorphic in nine or more species across both clades. The number of polymorphic loci in the species varied from five in *C. puniceus* to eleven in *D. s. capensis*. In the case of our targeted species, a total of 13 polymorphic markers were identified in the white steenbras (nine *de novo* developed and four through cross-species amplification) and 16 in the hottentot seabream (nine developed and seven through cross-species amplification).

2.3.5 Correlation between genetic distance and amplification success and polymorphism

Appendix 2E summarises the sequence divergence between the two source species (*L. lithognathus* and *P. blochii*) and the other species, as well as the predicted and observed amplification and polymorphism for the tested loci. These are plotted in Fig 2.3A, B, C & D. In each case the analysis only includes the markers developed in the source species i.e. either hottentot or white steenbras markers. The correlation between genetic distance and amplification success was found as non-significant for both *P. blochii* (*P*-value = 0.541; Fig. 2.3A) and *L. lithognathus* (*P*-value = 0.309; Fig. 2.3B). As well, no correlation between genetic distance and number of polymorphic loci was found for both species with a *P*-value of 0.064 (Fig. 2.3C) for the hottentot seabream and a *P*-value of 0.260 (Fig. 2.3D) for the white steenbras.

Table 2.2 Number of alleles (Na), Observed (H_O) and Expected (H_E) heterozygosity and F_{IS} (Weir & Cockerham, 1984) in eleven sparid species found around South Africa.

CLADE B																	
	A. berda			*D. s. capensis*			*D. c. hottentotus*			*L. lithognathus*			*R. globiceps*			*R. holubi*		
Locus	**N_A**	**H_O:H_E**	**F_{IS}**	**N_A**	**H_O:H_E**	**F_{IS}**	**N_A**	**H_O:H_E**	**F_{IS}**	**N_A**	**H_O:H_E**	**F_{IS}**	**N_A**	**H_O:H_E**	**F_{IS}**	**N_A**	**H_O:H_E**	**F_{IS}**
PBt001	1	M																
PBt003				15	0.66:0.90	0.295**	10	0.93:0.82	-0.103	14	0.87:0.89	0.052	7	0.68:0.79	0.167	11	0.75:0.88	0.186*
PBt007				10	0.81:0.86	0.093							9	0.63:0.85	0.300*			
PBt011				7	1.00:0.81	-0.151												
PBt012																		
PBt014				10	0.93:0.83	-0.093	7	0.81:0.75	-0.051				13	0.80:0.89	0.143			
PBt018				7	0.75:0.57	-0.268				3	0.43:0.47	0.103						
LLt002							7	0.73:0.78	0.102	10	0.93:0.85	-0.059	7	0.46:0.68	0.366*			
LLt005							1	M		5	0.56:0.68	0.211	3	0.42:0.58	0.333			
LLt006	7	0.92:0.75	-0.199	11	0.64:0.82	0.257*	13	0.81:0.89	0.122	8	0.75:0.83	0.128				8	0.33:0.84	0.629***
LLt007	2	0.20:0.32	0.419							8	0.87: 0.79	-0.074				4	0.68:0.61	-0.086
LLt011				2	0.18:0.16	-0.074	3	0.06:0.30	0.812***	11	0.87:0.85	0.007						
LLt014										8	0.75:0.82	0.124						
LLt020										10	0.93:0.87	-0.044						
LLtr004	3	0.31:0.35	0.148	9	0.87:0.82	-0.029				5	0.62:0.72	0.164	9	0.92:0.86	-0.030	3	0.50:0.53	0.091
PA535	9	0.87:0.85	0.014	10	0.75:0.86	0.163	2	0.00:0.11	1.00*				10	1.00:0.84	-0.148	11	1.00:0.88	-0.096
CL011	8	0.81:0.79	0.005	17	1.00:0.91	-0.064	10	0.87:0.83	-0.012	5	0.25:0.65	0.640***	8	0.68:0.81	0.191	7	0.81:0.68	-0.147
SA043																		
PB106	7	0.62:0.79	0.248*	9	0.81:0.79	0.010				9	0.81:0.83	0.058						

Table 2.2 continued

	CLADE B			CLADE A											
		P. blochii		*C. nufar*			*C. laticeps*			*C. puniceus*			*P. undulosus*		
Locus	**N_A**	**H_O:H_E**	**F_{IS}**	**N_A**	**H_O:H_E**	**F_{IS}**	**N_A**	**H_O:H_E**	**F_{IS}**	**N_A**	**H_O:H_E**	**F_{IS}**	**N_A**	**H_O:H_E**	**F_{IS}**
PBt001	9	0.87:0.78	-0.080												
PBt003	10	0.87:0.87	0.030	5	0.40:0.39	0.023	10	0.93:0.85	-0.059				12	0.62:0.88	0.324***
PBt007	11	0.87:0.89	0.050				5	0.31:0.41	0.279	10	0.33:0.84	0.627***	5	0.25:0.78	0.750**
PBt011	14	0.87:0.90	0.060				13	0.56:0.89	0.399***	13	0.75:0.86	0.165	12	0.93:0.87	-0.039
PBt012	18	0.81:0.93	0.159*												
PBt014	11	0.75:0.88	0.180												
PBt018	13	0.68:0.89	0.262**	10	0.50:0.78	0.400**							8	0.28:0.70	0.618***
LLt002				1	M		1	M		1	M				
LLt005	5	0.50:0.49	0.016				1	M		3	0.00:0.62	1.000*			
LLt006	10	0.81:0.85	0.085	1	M								1	M	
LLt007															
LLt011															
LLt014															
LLt020															
LLtr004	8	0.75:0.82	0.124	4	0.50:0.64	0.260	10	0.87:0.85	0.007				8	0.93:0.75	-0.142
PA535	19	0.87:0.89	0.056	8	0.81:0.82	0.049	20	0.87:0.93	0.093	16	0.66:0.92	0.310***	13	0.87:0.87	0.032
CL011	8	0.87:0.81	-0.037	5	0.68:0.61	-0.086	9	1.00:0.85	-0.132	15	0.87:0.90	0.067	5	0.68:0.59	-0.119
SA043	14	0.93:0.89	-0.011												
PB106	9	0.75:0.79	0.089	7	1.00:0.72	-0.352									

M = monomophic; **P*-value < 0.05, ***P*-value < 0.01 and ****P*-value < 0.001

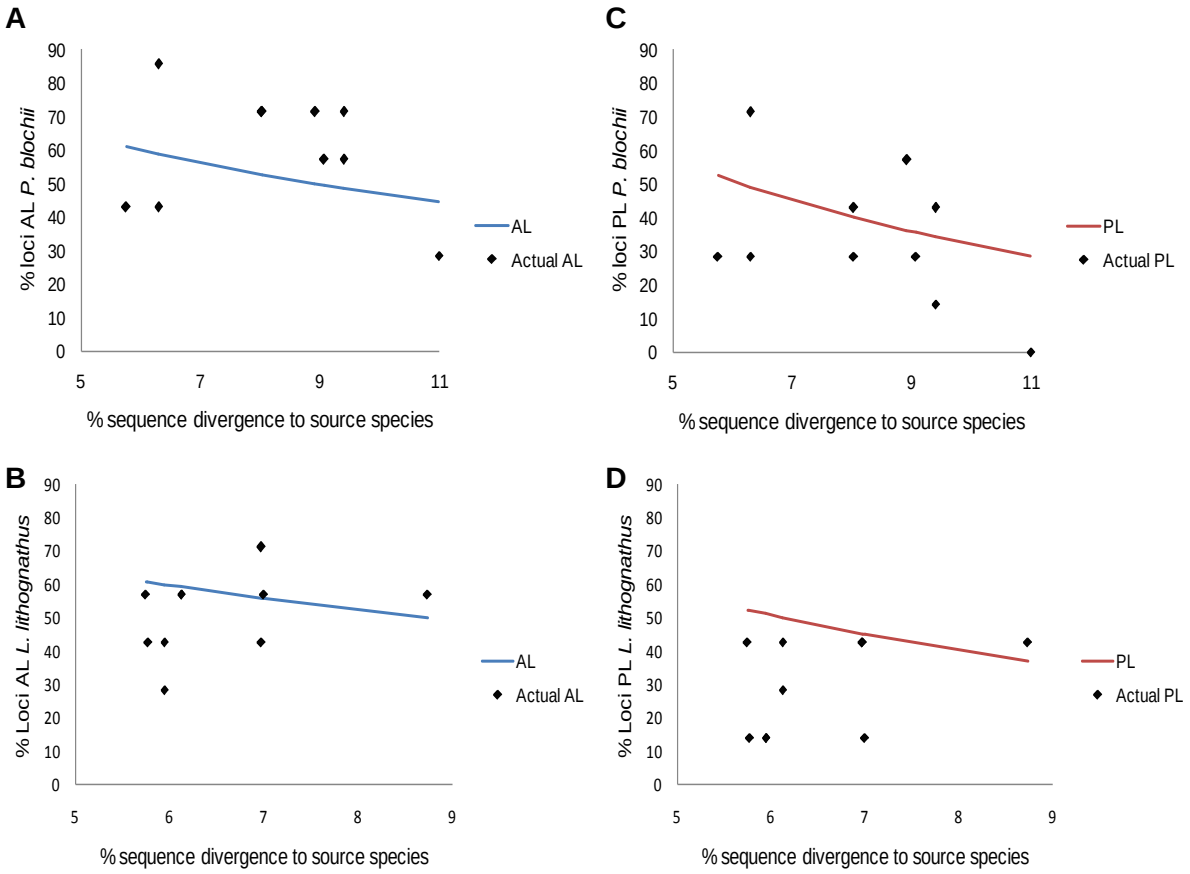


Fig 2.3 Relationship between the percentage sequence divergence between source species and target species and the percentage of amplification (AL) and polymorphism (PL) of the microsatellite loci. The trend lines represent the predicted negative correlation from the correlation equations of Carreras-Carbonell *et. al.* (2008) for percentage AL (blue) and PL (red) and the observed amplification success for **A**) *P. blochii* and **B**) *L. lithognathus* loci and polymorphism success **C**) *P. blochii* and **D**) *L. lithognathus* are shown (see Appendix 2E).

2.4 Discussion

In this study a comprehensive set of microsatellite markers were developed for the hottentot seabream and white steenbras through microsatellite development and cross-species amplification. The microsatellite markers were developed with a new approach combing the FIASCO protocol and 454 LifeScience/Roche sequencing. Although a number of recent studies have used 454 sequencing to identify sequences containing repeats (Abdelkrim *et al.* 2009; Allentoft *et al.* 2009; Santana *et al.* 2009, Castoe *et al.* 2010; Perry & Rowe 2010; Malausa *et al.* 2011), this is one of the first applications on marine fish.

These microsatellite markers were then successfully used as a panel to identify polymorphic markers in nine other economically important sparids. The genetic divergence of species included in this study from the source species (*P. blochii* and *L. lithognathus*) was then compared to the amplification success and the number of loci that were polymorphic. No significant negative correlation was identified in either species which is different to what has been found in general across a number of marine fish families in a recent study (Carreras-Carbonell *et al.* 2008).

2.4.1. *Development of microsatellites with 454 pyrosequencing*

Microsatellite markers were developed in the hottentot seabream and white steenbras using a repeat-enriched genomic fraction generated through the FIASCO protocol (Zane *et al.* 2002) and 454 sequencing. Through this protocol nine polymorphic microsatellite markers were identified in the hottentot (Table 2.1A) and nine polymorphic loci in the white steenbras (Table 2.1B).

This method allowed for the identification of a large portion of sequences containing repeats (4531 in *P. blochii* & 2839 *L. lithognathus*) and from this identification of a number of suitable loci for the hottentot seabream and white steenbras. In the case of the white steenbras, primers were designed for all potential tri- and tetranucleotide repeats with sufficient flanking region (40 potential loci). For the hottentot seabream, from the first twenty tetranucleotide sequences we identified nine polymorphic loci and due to time and cost constraints did not develop primers for the other potential loci with sufficient flanking region.

Using 454 LifeSciences/Roche GS-FLX sequencing instead of cloning is relatively time efficient and cost effective in terms of developing microsatellites as has been shown by recent studies (Santana *et al.* 2009; Csencsics *et al.* 2010). The number of sequences

generated is more than is feasible with cloning, which provides more options for choosing high quality sequences and higher success rates in isolating microsatellite repeat-containing loci. In the present study, when the cost per probing was divided by the number of sequences obtained through 454 sequencing in this study, it was found that the cost was approximately R0.50 per sequence compared to R 50 per forward and reverse sequence with Sanger sequencing (including BigDye reaction). This comparison demonstrates that using 454 sequencing to develop microsatellites is both time-efficient but also cost-effective (>100 times cheaper than Sanger sequencing).

A drawback to 454 sequencing is that often the sequences obtained do not have enough flanking region for primer design, which is observed less frequently with Sanger sequencing of cloned fragments. However, many sequences are obtained, thus there are still hundreds of sequences containing repeats for which primers may be designed. Additionally, genome walking can be used to obtain flanking regions when required.

There are some known sequencing discrepancies (insertions or deletions of false bases) which occur with 454 sequencing at A/T homopolymer runs (Moore *et al.* 2006). These regions should therefore be avoided in primer design. In the present study, owing to the relatively small fraction of sequencing for each species (only 1/5th of a 454 sequencing lane) there was generally only one-fold coverage of each repeat identified, which led to sequencing errors not being identified until the homozygotes were Sanger sequenced. A number of sequencing errors were observed between the 454 sequencing and Sanger sequencing which mainly consisted of insertions or deletions of A/T repeats (Appendix 2B & 2C). Forty three percent (7/16 loci) of the 454 sequences had sequencing errors when compared to the Sanger sequenced alleles. This may also influence how many primer pairs would need to be tested before the required number of markers is obtained.

Markers developed in the hottentot seabream were highly variable with an average of 17 alleles per locus compared to the white steenbras which had an average of 10 alleles per locus. This was confirmed by a Mann-Whitney test which found a significant difference (*P*-value = 0.002) between the mean number per locus for the hottentot and white steenbras. This lower variability may be due to smaller effective population size and/or population bottlenecks that resulted in a reduction in the allelic diversity in the white steenbras. However, the levels of heterozygosity are generally comparable across the two species and significant F_{IS} estimates appear to be due to the presence of null alleles at specific loci in both species. This could also be due to weak population genetic differentiation along the

distribution of the species, as the locations where these samples were collected are geographically separated.

The development of the markers in both species is important as both require population studies to identify if there are multiple stocks along the coast of South Africa. These markers were also developed with a view to being used to identify polymorphic loci in closely related species.

2.4.2. *Cross-species amplification and correlation testing*

Cross-species amplification has been successfully tested in a number of sparid studies (Brown *et al.* 2005; Chen *et al.* 2005; Piñera *et al.* 2006; Ponce *et al.* 2006; Castro *et al.* 2007; Liu *et al.* 2007; Piñera *et al.* 2007; Roques *et al.* 2007a; Roques *et al.* 2007b; Pérez *et al.* 2008). Through cross-species amplification of the microsatellite loci included in this study I was able to increase the number of microsatellite loci in each target species (13 loci white steenbras; 16 loci hottentot seabream). Also, a panel of 19 microsatellites were identified that are potentially useful in all the sparids.

A number of polymorphic loci were identified in the sparid species used in this study (Table 2.2) with a range of five loci in *C. puniceus* to eleven in *D. s. capensis*. Through this cross-species testing we were also able to increase the total number of loci for both the white steenbras (nine *de novo* developed and four cross-species) and the hottentot seabream (nine developed and seven through cross-species amplification).

When selecting markers through cross-species amplification it may be best to test loci from species more closely related i.e. from the same clade (Carreras-Carbonell *et al.* 2008). However, it was found that attempting to predict which source species would provide the most polymorphic loci in a species of interest may not be possible for this specific fish family. This is not the only sparid study to indicate that there is no significant negative correlation between genetic distance and amplification success or levels of polymorphism. In the study conducted by Carreras-Carbonell *et al.* (2008), in which a negative correlation across 10 marine fish families was found, there was no correlation when the sparid data were analysed on their own. The sparid information for this study was obtained from a cross-species study using *Sparus aurata* (Brown *et al.* 2005).

The fact that there was no significant negative correlation between genetic distance and amplification success or polymorphism in the sparids may be due to the small range of 16S genetic divergence (5 – 11 %) and the lack of very closely related species. It is observed in the study by Carreras-Carbonell *et al.* (2008) that high amplification success is only observed when genetic divergence of 16S is less than 5 %. However, for polymorphism according to the study by Carreras-Carbonell *et al.* (2008) a negative correlation should have been observed with the range of genetic divergence of 5-11 %; this was not observed in either the hottentot seabream (*P*-value of 0.064; Fig. 2.3C) or the white steenbras (*P*-value of 0.260; Fig. 2.3D).

A study by Liu *et al.* (2007) further indicated that it may be difficult to identify polymorphic loci in some sparids, as they identified 16 polymorphic markers for *Acanthopagrus schlegeli* after cross-species testing of 68 loci from six sparid species. This may be because the *Acanthopagrus* species is the most genetically divergent from all the other sparid species. Although it groups in clade B using 16S it is still the most genetically divergent species even to *P. blochii* and *L. lithognathus* which are also in clade B (see Appendix 2E; Chiba *et al.* 2009).

2.4.3. Conclusion

This study indicated the efficiency on both a time and cost level of using the FIASCO protocol with 454 sequencing to develop *de novo* microsatellites in non-model organisms. Some of the shortcomings of 454 sequencing were also indicated and discussed. Cross-species amplification of loci in the sparids was relatively effective although no relationship between genetic distance and microsatellite reliability (amplification and polymorphism) was observed on the scale of the family of this specific marine fish family.

The current study illustrates that whether isolating markers through *de novo* isolation or cross-species amplification, choosing the markers should be done with care. Sequencing a selection of alleles to identify that the variation is due to the repeat unit and not the flanking region will help to reduce high levels of homoplasy in genotypic data sets which may skew results. Also, general summary statistics should be calculated from relatively large sample sizes to identify the presence of null alleles. With this information the best markers can be chosen to decrease levels of homoplasy and null alleles.

The markers developed in this study will be useful tools for understanding the population genetic structure of the hottentot seabream and white steenbras, not only through the

markers developed within these species but also with additional markers identified through cross-species amplification. The polymorphic loci identified in the other sparids included in this study will also be useful in future studies aimed at elucidating the population genetic structure of these species.

2.5 References

Abdelkrim J, Roberson BC, Stanton JL, Gemmell NJ (2009) Fast, cost-effective devlopment of species-specific microsatellite markers by genomic sequencing. *BioTechniques* **46**, 185-192.

Allentoft ME, Schuster SC, Holdaway RN*, et al.* (2009) Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. *BioTechniques* **46**, 195-200.

Barbara T, Palma-Silva C, Paggi GM*, et al.* (2007) Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology* **16**, 3759-3767.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996-2004) GENETIX 4.05 Logciel sous Windows pour la Genetique des Populations, Laboratoire Genome, Populations, Interactions, CNRS UMR 5171, Universite de Montpellier II, Montpellier (France).

Bennett BA (1993) Aspects of the biology and life history of white steenbras *Lithognathus lithognathus* in southern Africa. *South African Journal of Marine Sciences* **13**, 83-96.

Broquet T, Mernard N, Petit E (2007) Non-invasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conservation Genetics* **8**, 249-260.

Brown RC, Tsalavouta M, Terzoglou V, Magoulas A, McAndrew BJ (2005) Additional microsatellites for *Sparus aurata* and cross-species amplification within the Sparidae family. *Molecular Ecology Notes* **5**, 605-607.

Buschiazzo E, Gemmell NJ (2006) The rise, fall and renaissance of microsatellites in eukaryotic genomes. *BioEssays* **28** 1040-1050.

Carreras-Carbonell J, Macpherson E, Pascual M (2008) Utility of pairwise mtDNA genetic distances for predicting cross-species microsatellite amplification and polymorphism success in fishes. *Conservation Genetics* **9**, 181-190.

Castoe TA, Poole AW, Gu W*, et al.* (2010) Rapid identification of thousands of copperhead snake (*Agkistrodon contortrix*) microsatellite loci from modest amounts of 454 shotgun genome sequence. *Molecular Ecology Resources* **10**, 341-347.

Castro J, Pino A, Hermida M*, et al.* (2007) A microsatellite marker tool for parentage assessment in gilthead seabream (*Sparus aurata*). *Aquaculture* **272**, 210-216.

Chen SL, Liu YG, Xu MY, Li J (2005) Isolation and characterization of polymorphic microsatellite loci from an EST-library of red sea bream (*Chrysophrys major*) and cross-species amplification. *Molecular Ecology Notes* **5**, 215-217.

Chiba SN, Iwatsuki Y, Yoshino T, Hanzawa N (2009) Comprehensive phylogeny of the family Sparidae (Perciformes: Teleostei) inferred from mitochondrial gene analyses. *Genes Genet. Syst* **84**, 153-170.

Csencsics D, Brodbeck S, Holderegger R (2010) Cost-effective, species-specific microsatellite development for the endangered dwarf bulrush (*Typha minima*) using next-generation sequencing. *Journal of Heredity* **101**, 789-793

Di Rienzo A, Peterson AC, Garza JC*, et al.* (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proc. natl. Acad. Sci. USA* **91**, 3166-3170.

Droege M, Hill, B (2008) The genome sequencer FLX^{TM} system- Longer reads, more applications, straightforward bioinformatics and more complex data sets. *Journal of Biotechnology* **136**, 3-10

Ellegren H (2004) Microsatellites: Simple sequences with complex evolution. *Nature Reviews: Genetics* **5**, 435-445.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver 3.11: An integrated software package for population genetics data analysis. *Evolution Bioinformatics* **1**, 47-50.

Faircloth BC (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources* **8**, 92-94.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98.

Heemstra P, Heemstra E (2004) *Coastal Fishes of Southern Africa* South African Institute for Aquatic biodiversity

Kruglyak S, Durett RT, Schug MD, Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from balance between slippage events and point mutations. *Proc. natl. Acad. Sci. USA* **95**, 10774-10778.

Liu YG, Liu LX, Wu ZX*, et al.* (2007) Isolation and characterization of polymorphic microsatellite loci in black sea bream (*Acanthopagrus schlegeli*) by cross-species amplification with six species of the Sparidae family. *Aquatic Living Resources* **20**, 257-262.

Malausa T, Gilles A, Maglécz E *et al.* (2011) High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources* Advanced publication

Moore MJ, Dhingra A, Soltis PS*, et al.* (2006) Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biology* **6**, 1-13.

Pérez L, Infante C, Ponce M*, et al.* (2008) Characterization of eight microsatellite markers in the white sea bream, *Diplodus sargus* (Teleostei, Sparidae). *Molecular Ecology Resources* **8**, 1291-1293.

Perry C, Rowe L (2011) Rapid microsatellite development for water striders by next-generation sequencing. *Journal of Heredity* **102**, 125-129

Piñera JA, Bernardo D, Blanco G, Vazquez E, Sanchez JA (2006) Isolation and characterization of polymorphic microsatellite markers in *Pagellus bogaraveo*, and

cross-species amplification in *Sparus aurata* and *Dicentrarchus labrax*. *Molecular Ecology Notes* **6**, 33-35.

Piñera JA, Bernardo D, Blanco G, Vazquez E, Sanchez JA (2007) Usefulness of microsatellite markers developed from *Pagellus bogaraveo* to genetically study five different species of Sparidae. *Marine Ecology* **28**, 184-187.

Ponce M, Infante C, Crespo A*, et al.* (2006) Characterization of microsatellite loci for the redbanded seabream, *Pagrus auriga* (Teloestei, Sparidae). *Molecular Ecology Notes* **6**, 527-529.

Primmer CR, Painter JN, Koskinen MT, Palo JU, Merila J (2005) Factors affecting avian cross-species microsatellite amplification. *Journal of Avian Biology* **36**, 348-360.

Rico C, Rico I, Hewitt G (1996) 470 Million years of Conservation of Microsatellite Loci among Fish Species. *Proceedings of the Royal Society London: Biological Sciences* **263**, 549-557.

Roques S, Galarza JA, Macpherson E*, et al.* (2007a) Isolation of eight microsatellite loci from the saddled bream *Oblada melanura* and cross-species amplification in two sea bream species of the genus *Diplodus*. *Conservation Genetics* **8**, 1255-1257.

Roques S, Galarza JA, Macpherson E, Turner GF, Rico C (2007b) Isolation and characterization of nine polymorphic microsatellite markers in the two-banded sea bream (*Diplodus vulgaris*) and cross-species amplification in the white sea bream (*Diplodus sargus*) and the saddled bream (*Oblada melanura*). *Molecular Ecology Notes* **7**, 661-663.

Rousset F (2008) Genepop'007: A complete reimplementation of GenePop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Rozen S, Skaletsky H (2000) PRIMER 3 on the WWW for general users and for biologist programmers. *Bioinformatics methods and protocols: Methods in molecular biology*, 365-386.

Santana QC, Coetzee MPA, Steenkamp ET*, et al.* (2009) Microsatellite discovery by deep sequencing enriched genomic libraries. *BioTechniques* **46**, 217-223.

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**, 615-629.

Soulsbury CD, Iossa G, Edwards KJ (2009) The influence of evolutionary distance between cross-species microsatellites and primer base-pair composition on allelic dropout rates. *Conservation Genetics* **10**, 792-802.

Swofford DL (2001) PAUP*, Phylogenetic Analysis Using Parsimony (*and other methods) version 4. *Sinuaer Associates, Sunderland, Mass*

Teske PR, Cowley PD, Forget FR, Beheregaray LB (2009) Microsatellite markers for the roman, *Chrysoblephus laticeps* (Teleostei: Sparidae), an overexploited seabream from South Africa. *Molecular Ecology Resources* **9**, 1162-1164.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acid Research* **25**, 4876-4882.

Van der Elst R (1988) *A guide to the common sea fishes of Southern Africa*, 2nd edn. C. Struik Publishers.

van Oosterhout C, Hutchinson WF, Wills DP, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.

Wattier R, Engel CR, Saumitou-Laprade P, Valero M (1998) Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Molecular Ecology* **7**, 1569-1573.

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology* **11**, 1-16.

2.6 Appendices

Appendix 2A Species included for the determination of sequence divergence between sparid genera based on mitochondrial DNA 16S rDNA sequences.

Species	GenBank accession number
Pachymetopon aneum	AF247400
Lithognathus mormyrus	AF247410
Cheimerius nufar	AF247405
Diplodus cervinus	AF247420
Chrysoblephus cristiceps	AF247417
Polysteganus praeorbitalis	AF247408
Rhabdosargus thorpei	AF247429
Acanthopagrus berda	AF247413

Appendix 2B Sequences obtained from 454 sequencing of microsatellite loci developed for (**A**) *P. blochii* and (**B**) *L. lithognathus*. Repeat sequences and primer binding sites are indicated in bold.

PBt001
AGGTTCCCCACAGAAGGTCTTTCTGTGTATTATAACCCATCTATTTCCTCAACCCTGATAAATGGAAATATGGAGACTTGTAGTCCAT**AGATAGATAGATAGATAGATAGATAGATA GATAGAT**TCTTGTACTTGAAATTTCCTTTCAATGTGAAATCCAATTATAGACATGACCTTTGCCCCGTAAAACATACAAATGCAGTTA**GGGCTGTTTCCCAGGTACT**

PBt003
TTGGGGAAAAGGAGAGGCGAAAATGTGCAAGGTTATGGCATGCAGCAATACTTAGGG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAG ATAGATAGATA**GCACACAGACGGCAAAAAT**CAACATACAGAGCGAGTCGG**

PBt006
CATTTTGGCAGTCCGTCCGAGGACGAAGGTTTCTTTGAAAGGATTTCCTAACATTGCAATTTCAGGTGTCTGTAGCCATTTTGCTGCATTTAGATCCAGAACTAACTTTGCAAAAGAA ACAATCATAGATAGATATAGCAGGC**AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT**ACCGCACTATTATGTATGACTTCACTGTATTCAT**GGATACGAACGTC AGTACATTT**

PBt007
CGCAGGATTCATACGTAAAGAAATGATACACACAACTGTGATAATC**TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC**TGTCTGTCTATGGATCTATAG TGTTGTAGGTGGTCTGCAT**GGCAGGGCATGGAATGGAA**

PBt011
CACAGACGGGAATGCACACAGAACCTTCAGTTCACCGACAATCACTCTCTTATCTTTCCACTCCTC**TATCTATCTATCTGTCTATCTATCTATCTATCTATCTATCTATCTATCTATC TATCTATCTATCTATC**TATTTATCACTCTTTGTCTCGCTGTCTATCTCAAAGAGCAATGCATGGCCTATCACTTTTCCTCTCTGA**CGTTCTGTCCTGACACCCT**

PBt012
CACGCACTGGTATCGAATCTGAACTTGGTATCGGCCGGTACCCCAAGCCTGGAATCAGTATCGAGAAAGGAAAAAAATGGTATCGGAACATCTCTTTCATATCA**TATCTATCTATCT ATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC**CAACACTCAATT**GACAGACTAGCTCCACCTCT**

PBt013
GAGAAACAACAGTGTCAGGGGGAAGTACAGTGTCATA**GATAGATAGATAGACAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**AATAGGAAA GGTCTAACTGTTTGTTTTGT**CCTGTGTGTCATCACTCTTGC**

PBt014
GTTGGGTGATGGAGCCAAGATAACTAGATTGATAGATACTTCAGGTCAATCGATACTTCAGGTCGATC**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT AGATAGATAGATAGATAGATAGATA**GATTAGATAACGAT**GGGGCTTCTAGTCACCTCTG**

PBt018
GTCAAGCCTGTCAAACTGGGTAGTGATCCTG**TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC**CTGTAAACTATGATCGACACAAAAACAATTTC ACTGTCGACTTTCTGTTTTTCTACTTCTAAAAAAGTTGGTCCAGTCGGGTTCCTCTGTATAGCCTGAATGTTCCTCCGA**CGGCTCGTGTTTCTTTCCA**
(A)

LLt002:
TAACACAGGGCGACTGGGTTACAAGACT**GCTTTGACCGTCTTTTCCCC**ACAGAGAAACATTTGGAATATTTATGTAGATCTTTTTT**CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT**CA
GTAGCTGAATTATTCTGAATCAGCATGGCTGCGACTAAGGCTAATTTTCACTATCTGTTGATTTATCATTTTATCTATGAAATATCACAAAATAGTGACAAGCGGCGAGAGTCTTTAGA
TGTCTTTTTTTGA**CCGACCAACAGCCCAAAAC**

LLt005
TAACGAAA**GTGAGAATCCTCCATCCAAGTC**TCATGGAAATCAGTGCAGTAGTTTTTGTGTAAGCCTGCTGAGAAGCCAA**CAAACAAACAAACAAACAAACAAACAAACAAACAAA**
CAAACAAACGGACGGACGCAGGTGAAAACAGGTCAATGTTACAAGGTTACAAATCACAATTGAGCACATTTTTCTTCAATTCAATGAAAGAATATCTAAATATAAAAATGG**CAGCAT**
GGAAAGGCCCATT

LLt006
TCATCTGACCACAGACCAGCTATTGTACAAAATTACTGAGCTGATATATATCATAAATGAAGCAATAAACAATTTTTACTCTTTGACTTTGAAGCTGCTGTAGGCACATTTTGACTGAC
AGGCTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**GGAATGTAAACACTACTGTC**CTTACTGTCACAAATGGCGTG**

LLt007
TGGATACCTGAATGTCCCTCGGGATAAATAAAGTATCTGTC**TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC**TGTGACCTTGAAGAGGCAGGCCATA
CACGATAAATAGCAATAAAAAGAAACATCAGAGTGAAGAGGCATCAAATTGAGATTTCATGATCCA**GTGGAGCCCAGGGTATGAC**

LLt011
AAACCCATCACCCAGTCCCATCAACGCCCCCCAACCTCATTTACAGTCAGATCCCGTCAGGTCCCGACCCCTGCCGACACTGTAACTGTGGCAACCAGGAAGATGCAGAGAGACA
GTAAAGAGTCGCAAAATGGCGAGGA**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**GAATGGTCTGTTCAACCAAAATCAAAAATA**CAGCGTCTAATC**
TGTAGTGTTTT

LLt014
ATACAAATGCCACTAGACCAACAACTAGAAAAGAAATGCTCAGAACTAGATAAATACATGACAGAAGAGATAGATAGATA**AGATAGATAGATAGATAGATAGATAGATAGATAGAT**
AGATAGATTAGAGTTTAGGA**CATCACTGACTACCAATCTTACTATAG**

LLt020
CTGTGCCTTTATCTGTATCTACTCTGGGCCGGGACCTAACCACCATCCTAGTTTGGTGGAAATGCATTCAGTAGTTTTGTGTAATCCTGTTGAC**CAACCAACCAACCAACCAACCA**
ACCAACCAACCAACCAACCAACCAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAATGCCCAGGGCCAATTACAGCTCTCCATACATGACATGAA
GTAGGGTTGGGTAATTTCTGCCAAATTAGCATATGACGATGTCTATAGT**GAAACATCGTGATGGACAA**

LLt024
TGACAGAATGAACGTCGGCAATATGTTTCTGCAAACAGAGGTTATGTTGCAACGTCACGTTCCTTCAGTTTCACATTTTACGTTGTGACAACGCTGTTCTAATGGTCTTATGGTTTAG
TTGGTTAGTT**AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT**AGCAGGAGCTGGGANAAAT**CCAGACCTCAGAA**
ACAGAACC

LLtr004
CGCTCATTGATTGGAGCCCATTGTAATAGCAATGGAAATGTATTCAGATAATTATCAGGCTCATTTAGGTATGTGGCTTGGTGTTGCTTTTGCGAAGGCAGACAGAGAGATTTTTTTC
TCCCG**CAGCAGCAGCAGCAGCAGCAGCAGCAGCAG**AGCCTCAGAGCTGACAGTAAATTGCTCAT**CAACCAGTTTCGTTCCAGTGT**
(B)

Appendix 2C Homozygotes sequenced to assess repeat and flanking region variability of alleles at microsatellite loci of (**A**) *P. blochii* and (**B**) *L. lithognathus*, compared with the allele length from fragment analysis (shown in parentheses on the left of each sequence); DNA sequence and number of repeats are provided for each sequenced allele.

PBt001
(225) GATAATGGAAATATGGAGACTTGTAGTCCAT**AGATAGATAGATAGATAGATAGATAGATAGATAGAT--------**TCTTGGACTTGAAATTTCCTTTCAATGTGAAATTC : 9
(229) GATAATGGAAATATGGAGACTTGTAGTCCAT**AGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT**TCTTGGACTTGAAATTTCCTTTCAATGTGAAATTC : 10

PBt003
(183) TTAGGGGAA**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA--------**GCACA : 21
(187) TTAGGGGAA**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**GCACA : 22

PBt007
(154) GA**TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC--------**TGTCTGTCTATGGATCTATAGTGTTGTAGGTGGTCTGCATGGCAG : 16
(158) GA**TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC**TGTCTGTCTATGGATCTATAGTGTTGTCGGTGGTCTGCATGGCAG : 17

PBt011
(191) TCC**TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA----------**TTTATCACTCTTTGTCTCGCTGTCTATCTCAAAG : 11
(215) TCC**TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA--------**TTTATCACTCTTTGTCTCGCTGTCTATCTCAAAG : 17
(219) TCC**TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA**TTTATCACTCTTTGTCTCGCTGTCTATCTCAAAG : 18

PBt012
(164) AAAAAATGGTATCGGAACATCTCTTTCATATCA**------TATCTATCTATCTATCTATCTATCTATC----------**CA : 7
(176) AAAAA--TGGTATCGGAACATCTCTTTTATTTAAACG**TATCTATCTATCTATCTATCTATCTATCTATCTATC----------**CA : 9
(180) AAAAA--TGGTATCGGAACATCTCTTTCA**----------TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC----------**CA : 12
(188) AAAAA--TGGTATCGGAACATCTCTTTCATATCA**------TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC----------**CA : 13
(204) AAAAAATGGTATCGGAACATCTCTTTCATATCA**------TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC**CA : 17

PBt013
(126) TCAGAT**GATAGATAGACAGATAGATAGATAGATAGATAGATA----------**AATAGGAAAGGTCTAACTGTTTTGTTTTTGTCCTGTGTGTCATCACTCTTGCAA : 9
(134) TCAGAT**GATAGATAGACAGATAGATAGATAGATAGATAGATAGATA**AATAGGAAAGGTCTAACTGTTTTGTTTTTGTCCTGTGTGTCATCACTCTTGCAA : 11

PBt014
(161) TACTTCAGGTCGATC**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**----------GAATTAGATAAACGA : 14
(177) TACTTCAGGTCGATC**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**GAATTAGATAAACGA : 18

PBt018
(188) CCT**CTATCTATCTATCTAT----------**CCTGTAAACTATGATCGACACAAAAACAATTTCACTGTCGACTTCTGTTTTCT : 4
(196) CCT**CTATCTATCTATCTATCTATCTAT----------**CCTGTAAACTATGATCGACACAAAAACAATTTCACTGTCGACTTCTGTTTTCT : 6
(212) CCT**CTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT**CCTGTAAACTATGATCGACACAAAAACAATTTCACTGTCGACTTCTGTTTTCT : 10
(A)

LLt002
(258) TAGATCTTTTTT**CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT**CAGTAGCTGAATTATTCTGAATCAGCATGGCTGCAACTAAGGCT : 12

LLt005
(219) AGCCTGCTGAGAAGCCAA**CAAACAAACAAACAAACAAACAAACAAACAAA**----------CGGACGGACGCAGGTGAAAACAGGTCAATGTTACAAG : 8
(223) AGCCTGCTGAGAAGCCAA**CAAACAAACAAACAAACAAACAAACAAACAAACAAA**---------CGGACGGACGCAGGTGAAAACAGGTCAATGTTACAAG : 9
(227) AGCCTGCTGAGAAGCCAA**CAAACAAACAAACAAACAAACAAACAAACAAACAAACAAA**CGGACGGACGCAGGTGAAAACAGGTCAATGTTACAAG : 10

LLt006
(209) GACAGGCTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**----------GGAATGTAAACACTACTGTCCTTACTG : 11
(213) GACAGGCTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**----------GGAATGTAAACACTACTGTCCTTACTG : 12
(225) GACAGGCTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**GGAATGTAAACACTACTGTCCTTACTG : 15
(225) GACAGGCTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**GGAATGTAAACACTACTGTCCTTACTG : 15

LLt007
(189) TGTC**TATCTATCTATCTATCTATCTATCTATCTATCTATC**----------TGTGACCTTGAAGAGGCAGGCCATACACGATAAATAGCAATAAAAAGAAAAC : 10
(197) TGTC**TATCTATCTATCTATCTATCTATCTATCTATCTATCTATC**TGTGACCTTGAAGAGGCAGGCCATACACGATAAATAGCAATAAAAAGAAAAC : 12

LLt011
(251) TAAAGAGTTGCAAATGGCGAGGA**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**--------GAATGGTCTGTTCAACCCAA : 14
(255) TAAAGAGTTGCAAATGGCGAGGA**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**GAATGGTCTGTTCAACCCAA : 15

LLt014
(148) GACAGAAGAAAT**AGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT**----------AGAGTTTAGGACAT : 10
(164) GACAGAAGAAAT**AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT**AGAGTTTAGGACAT : 14

LLt020
(274) TCCTGCTGAC**CAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAACAAACAAACAAACAAACAAACAAACAAACAAA**----------TGCCC : 19
(282) TCCTGCTGAC**CAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAACAAACAAACAAACAAACAAACAAACAAACAAACAAA**TGCCC : 21

LLtr004
(198) TGCTTTTGCGAAGGCAGATAGAGAGATTTTTTTCTCCCG**CAGCAGCAGCAGCAGCAGCAGCAG**----------AGCCTCAGAGCTGACAGTAAATTTGCTCATCAAC : 8
(201) TGCTTTTGCGAAGGCAGATAGAGAGATTTTTTTCTCCCG**CAGCAGCAGCAGCAGCAGCAGCAG**-------AGCCTCAGAGCTGACAGTAAATTTGCTCATCAAC : 9
(204) TGCTTTTGCGAAGGCAGATAGAGAGATTTTTTTCTCCCG**CAGCAGCAGCAGCAGCAGCAGCAG**AGCCTCAGAGCTGACAGTAAATTTGCTCATCAAC : 10
(B)

Appendix 2D Summary of cross-species analysis for 19 microsatellite markers amplified in 11 sparid species. Amplification, size-range and percent null alleles (Oosterhout estimate) are recorded for each locus

CLADE B																	
	A. Berda			*D. s. capensis*			*D. c. hottentotus*			*L. lithognathus*			*R. globiceps*			*R. holubi*		
Locus	**A**	**Size range**	**% null alleles**	**A**	**Size range**	**% null alleles**	**A**	**Size range**	**% null alleles**	**A**	**Size range**	**% null alleles**	**A**	**Size range**	**% null alleles**	**A**	**Size range**	**% null alleles**
PBt001	N	-	-	Y	388	-	N	-	-	N	-	-	I	-	-	N	-	-
PBt003	N	-	-	Y	124-256	13.43	Y	128-172	-	Y	167-251	-	Y	129-165	-	Y	139-179	-
PBt007	N	-	-	Y	156-196	-	N	-	-	N	-	-	Y	119-167	12.62	N	-	-
PBt011	I	-	-	Y	195-227(St)	-	Y	St	-	N	-	-	N	-	-	Y	St	-
PBt012	I	-	-	N	-	-	N	-	-	N	-	-	N	-	-	I	-	-
PBt014	N	-	-	Y	141-177	-	Y	176-204	-	N	-	-	Y	145-221	-	I	-	-
PBt018	N	-	-	Y	184-224	-	N	-	-	Y	188-196	-	N	-	-	Y	St	-
LLt002	I	-	-	N	-	-	Y	230-266	-	Y	242-282	-	Y	169-241	14.97	N	-	-
LLt005	Y	St	-	I	-	-	Y	128	-	Y	219-223	-	Y	191-199	-	N	-	-
LLt006	Y	234-262	-	Y	368-420	-	Y	358-442	-	Y	205-237	-	N	-	-	Y	420-452	29.43
LLt007	Y	216-220	-	N		-	N	-	-	Y	185-229	-	I	-	-	Y	196-208	-
LLt011	N	-	-	Y	267-271	-	Y	234-254	30.26	Y	227-271	-	I	-	-	N	-	-
LLt014	N	-	-	N	-	-	N	-	-	Y	140-172	-	N	-	-	N	-	-
LLt020	N	-	-	N	-	-	N	-	-	Y	234-306	-	N	-	-	N	-	-
LLtr004	Y	199-253	-	Y	193-229	-	N	-	-	Y	198-210	-	Y	190-217	-	Y	181-187	-
PA535	Y	117-157	-	Y	127-171	-	Y	93-105	-	N	-	-	Y	132-180	-	Y	115-163	-
CL011	Y	228-256	-	Y	248-316	-	Y	230-266	-	Y	218-242	28.61	Y	180-208	-	Y	213-253	-
SA043	N	-	-	N		-	N	-	-	Y	St	-	N	-	-	N	-	-
PB106	Y	318-342	-	Y	243-335	-	Y	297-301	-	Y	289-321	-	N	-	-	N	-	-

N = no amplification, Y = amplification, I = inconsistent amplification, St = stutter

Appendix 2D continued

	CLADE B			CLADE A										
		P. blochii		*C. nufar*			*C. laticeps*			*C. puniceus*			*P. undulosus*	
Locus	A	Size range	% null alleles	A	Size range	% null alleles	A	Size range	% null alleles	A	Size range	% null alleles	A	Size range	% null alleles
PBt001	Y	209-253	-	N	-	-	Y	297	-	I	-	-	I	-	-
PBt003	Y	143-183	-	Y	216-276	-	Y	150-194	-	N	-	-	Y	165-221	14.34
PBt007	Y	130-170	-	I	-	-	Y	117-133	-	Y	121-177	29.77	Y	120-136	33.01
PBt011	Y	187-259	-	I	St	-	Y	160-220	18.71	Y	184-260	-	Y	186-242	-
PBt012	Y	144-256	-	N	-	-	Y	-	-	Y	St	-	N	-	-
PBt014	Y	153-197	-	N	-	-	Y	St	-	Y	St	-	N	-	-
PBt018	Y	184-284	11.65	Y	172-268	17.32	N	-	-	N	-	-	Y	136-264	27.74
LLt002	N	-	-	Y	201	-	Y	218	-	Y	218	-	N	-	-
LLt005	Y	191-215	-	N	-	-	Y	194	-	Y	194-226	43.92	N	-	-
LLt006	Y	376-424	-	Y	221	-	N	-	-	N	-	-	N	-	-
LLt007	N	-	-	I	-	-	N	-	-	N	-	-	Y	218	-
LLt011	N	-	-	N	-	-	N	-	-	N	-	-	N	-	-
LLt014	N	-	-	N	-	-	N	-	-	Y		-	N	-	-
LLt020	N	-	-	N	-	-	N	-	-	N	-	-	N	-	-
LLtr004	Y	199-220	-	Y	211-220	-	Y	216-249	-	N	-	-	Y	207-231	-
PA535	Y	137-249	-	Y	101-153	-	Y	113-249	-	Y	115-207	13.84	Y	103-155	-
CL011	Y	222-254	-	Y	241-281	-	Y	227-259	-	Y	239-303	-	Y	215-231	-
SA043	Y	182-274	-	N	-	-	N	-	-	N	-	-	N	-	-
PB106	Y	314-346	-	Y	311-335	-	N	-	-	N	-	-	Y	-	-

Appendix 2E Summary of the pairwise 16S sequence divergence (GD in %) estimated and predicted and observed amplification (AL in %) and polymorphism (PL in %) for the sparid genera with *P. blochii* and *L. lithognathus* as source species respectively.

Pachymetopon blochii

Species	GD	AL (predicted)	AL (observed)	PL (predicted)	PL (observed)
L. lithognathus	5.74	60.94	42.8	52.39	28.5
D. c. hottentotus	6.28	58.64	42.85	49.08	28.5
D. s. capensis	6.28	58.64	85.7	49.08	71.4
C. laticeps	8.02	52.42	71.4	40.11	42.8
C. puniceus	8.02	52.42	71.4	40.11	28.5
P. undulosus	8.9	49.75	71.4	36.25	57.1
C. nufar	9.06	49.31	57.1	35.63	28.5
R. globiceps	9.39	48.4	57.1	34.31	42.8
R. holubi	9.39	48.4	71.4	34.31	14.2
A. berda	10.98	44.39	28.5	28.53	0

Lithognathus lithognathus

Species	GD	AL (predicted)	AL (observed)	PL (predicted)	PL (observed)
P. blochii	5.74	60.94	57.1	52.39	42.8
C. laticeps	5.76	60.86	42.8	52.28	14.2
C. puniceus	5.93	60.86	42.8	52.28	14.2
P. undulosus	5.93	60.1	28.5	51.18	14.2
D. s. capensis	6.11	59.34	57.1	50.09	42.8
D. c. hottentotus	6.11	59.347	57.1	50.09	28.5
R. globiceps	6.96	56.025	71.4	45.3	42.8
R. holubi	6.96	56.025	42.8	45.3	42.8
C. nufar	6.99	55.93	57.1	45.17	14.2
A. berda	8.72	50.27	57.1	37	42.8

Chapter 3

Genetic differentiation in the hottentot seabream *Pachymetopon blochii* (Valenciennes, 1830) from South Africa based on microsatellite variation

Abstract

The hottentot seabream (*Pachymetopon blochii*) is an endemic, demersal sparid occurring along the west coast of southern Africa, found mainly in kelp beds and rocky outcrops. This species is targeted by line fisheries and is considered vulnerable due to its slow growth and sedentary adults. Two hypotheses are considered in this study, the first being that the eggs and larvae remain in the habitats where the adults are found and this would lead to less gene flow between geographically separated habitats (hypothesis of isolation). Alternatively, the eggs and larvae could be transported by the inshore currents which would lead to a single population identified along the west coast of South Africa (hypothesis of panmixia). The main aim of this study was to investigate which of these two hypotheses best explains the population connectivity in the hottentot seabream. For this purpose, the spatial and temporal genetic variation of this species along the South African west coast was assessed. Fourteen highly polymorphic loci were genotyped for 288 individuals across nine locations sampled in 2001 and in 189 samples from six locations in 2009. Weak spatial structure was identified between the sampling locations from 2009 using Factorial Correspondence Analysis (FCA), Analysis of Molecular Variance (AMOVA) and Spatial Autocorrelation (SAC). Between the 2001 sampling locations no significant spatial structure was identified. Temporal variation was identified between the two sampling years. This was likely due to “random genetic patchiness” which led to variations in the observed population structure across different years. In conclusion, weak spatial and temporal variation was observed between sampling years, which is likely due to the larval dispersal and mortality mediated through the oceanographic features along the west coast of South Africa.

Keywords: gene flow, hottentot seabream, *Pachymetopon blochii*, population structure, South Africa, Sparidae, spatial and temporal variation

3.1 Introduction

Population genetics in the marine environment mainly focuses on the identification of spatial and temporal genetic variation influenced by life-history traits and the habitat of the species of interest (White *et al.* 2010). It is important to identify the patterns of dispersal of pelagic larvae and adults as these traits influence genetic connectivity of populations (Hogan *et al.* 2010). The levels of connectivity (gene flow) between populations distributed in fragmented habitats in turn influence the metapopulation dynamics of the species and evolution of populations on an evolutionary and ecological scale (Hogan *et al.* 2010; Planes & Lemer 2011).

In the past, the ocean was perceived as an endless resource with many marine fish species having infinitely large populations with high levels of gene flow due to adult and larval dispersal (Waples 1998). These marine species have life-history characteristics that contribute to high gene flow such as adult dispersal, egg and larval drift, high fecundity and specific spawning areas which allow for large population sizes and extensive mixing (Waples 1998). These species have also developed an adaptive strategy that consists of extended larval periods allowing them to fully benefit from transient habitat features (ocean currents). Although a number of fish studies have identified high connectivity and panmixia (Roberts & Ayre 2010; White *et al.* 2010), there has also been records of some species showing genetic differentiation between populations (De Innocentiis *et al.* 2004; Gonzalez *et al.* 2008; Nielsen *et al.* 2009).

Although many strategies allow fish species to maintain high levels of gene flow, there are other features, which may lead to genetic isolation (Waples 1998). Stable oceanographic features between certain areas can be barriers to gene flow created by temperature, salinity, powerful currents or upwelling zones (Hauser & Carvalho 2008). Other factors can be historical, as populations of the same species may have been separated by habitat fragmentation linked to past sea level fluctuations (Patarnello *et al.* 2007). Genetic divergence between populations has been identified from ocean basin (White *et al.* 2010) to local scales (Saillant *et al.* 2010); this is likely due to a combination of life-history traits, oceanographic features and historical factors. To identify which of these factors best explains the genetic variation in a species, it is of major importance to understand its evolution.

Population genetic structure is considered a continuum from highly divergent multiple populations to a single panmictic population (see Fig 1.1; Waples & Gaggiotti 2006; Hauser & Carvalho 2008). The goal of marine population genetics investigations is to identify the

scale of genetic organisation of a specific species along this continuum and this, in turn, will indicate the level of connectivity between localities. In the marine environment the high levels of connectivity and large population sizes lead to low genetic differentiation. Ward *et al.* (1994) determined the median amount of genetic differentiation (F_{ST}) typically identified in marine fish species at about 0.02, which is considered low compared to freshwater fish (mean F_{ST} = 0.144). However, these low levels of differentiation may still be biologically significant (Waples 1998). Also it has been shown in the marine environment that an F_{ST} of 0.002 may indicate populations with biological relevance (Hauser & Carvalho 2008).

Marine organisms have high fecundity and often only a small number of mature adults will contribute to the next generation (Appleyard & Ward 2006). This, coupled with high mortality of eggs and larvae leads to a low effective population size (N_e), which is often two to five orders of magnitude lower than the actual census population size (N; Hauser & Carvalho 2008). The census population size shapes demographic and ecological processes, while the N_e determines how the population will respond to evolutionary factors (Waples 2010). The smaller the effective population size the stronger the effect of evolutionary factors such as genetic drift, mutation, selection and migration (Waples 2010). The underlying factor contributing to low N_e in marine fish is usually variance in reproduction success (Hauser & Carvalho 2008).

To determine population genetic structure in the marine realm the most common markers used are microsatellite markers (Selkoe & Toonen 2006). These co-dominant and highly polymorphic markers have allowed for many questions about the demography of marine organisms and the marine environment to be addressed. They are highly valuable for studying species with high gene flow, though in such species it is often difficult to detect underlying structure using classic population genetic approaches based on allelic frequency due to the low power and large population sizes (Waples & Gaggiotti 2006; Hauser & Carvalho 2008). Recently, a number of analyses based on individual genotypes combined with geographic data have been developed to address this problem (Kelly *et al.* 2010). An example of this is spatial autocorrelation (SAC; Peakall & Smouse 2006). These individual based analyses are interpreted in terms of the life-history and geographic distribution of species and along the South African west coast there are a number of oceanographic features (upwelling zones, currents and eddies) which may enhance or restrict gene flow.

The west coast of South Africa is characterized by cold nutrient rich water which is brought in a northerly direction from the Antarctic by the Benguela current (Shillington *et al.* 2006; Hutchings *et al.* 2009). This region has low species diversity but the species that occur in the

region are highly abundant (for example sardine and anchovy). There is a “jet” current moving in a northerly direction along the west coast which transports egg and larvae from the Agulhas Bank and then feeds into an inshore current which moves in a southerly direction along the coastline depositing the larvae in inshore habitats (refer to Fig. 1.2; Hutchings *et al.* 2009). Along the coast there are a number of small upwelling zones around Lamberts Bay, Paternoster, Saldanha Bay and Cape Point. There is a large upwelling zone around Lüderitz in southern Namibia which is a known barrier to gene flow for some marine organisms such as deepwater hake (*Merluccius paradoxus*; von der Heyden *et al.* 2007). A large portion of the west coast line fishery is made up of species of the demersal Sparidae.

The Sparidae is an economically important fish family around the world and of approximately 115 known species, 110 have been identified around South Africa, with 24 species being endemic (Heemstra & Heemstra 2004; Chiba *et al.* 2009). In the Sparidae some studies across the world have identified high connectivity and gene flow leading to the conclusion of panmixia (Arculeo *et al.* 2003; Roberts & Ayre 2010; Teske *et al.* 2010), while others have identified weak spatial and temporal structure (Planes & Lenfant 2002; De Innocentiis *et al.* 2004; Rossi *et al.* 2006; Burridge & Versace 2007; González-Wangüemert *et al.* 2007; González-Wangüemert *et al.* 2010). Along the coast of South Africa there is currently only one published population genetics study based on microsatellites in a sparid namely the roman (*Chrysoblephus laticeps*) for which one population with high connectivity was identified along the east coast from False Bay to Port Alfred (Teske *et al.* 2010).

The hottentot seabream is endemic to the west coast of Africa with a relatively patchy distribution from Cape Agulhas to Angola (Van der Elst 1988; Heemstra & Heemstra 2004). The hottentot is, however, most abundant along the South African coastline (Pulfrich & Griffiths 1988a; Punt *et al.* 1996). This species mainly occupies kelp beds and rocky outcrops which represent a relatively patchy habitat to a depth of about 50 m. Spawning occurs multiple times during the year with increased spawning activity in late autumn and summer. There is no specific spawning areas along the coast and it is thought that the adults spawn within their inshore habitats (Pulfrich & Griffiths 1988b). The eggs and larvae are thought to be pelagic; larval duration is not known and the juveniles and adults are found in the same habitats (Pulfrich & Griffiths 1988b). The adults do not migrate and are considered relatively sedentary. A previous mark-recapture study which tagged 3661 hottentot only re-sampled seven individuals, six in the same locations they were originally tagged in and one being caught 722 days later, 40 km away from its original capture location (Nepgen 1977). Approximately 600 tons of hottentot are caught annually off the west and east coast of the Western and Eastern Cape Provinces combined and this is the second most important line

fish after snoek (*Thyrsites atun*; ~ 6000 tons) along the west coast of South Africa (Sauer *et al.* 2003).

The hottentot seabream is a species of interest because of its importance to the west coast line fishery of South Africa and its life-history traits such as late sexual maturity, longevity and restricted adult movement that make it vulnerable to overexploitation. These life-history traits may also have led to population differentiation. It is important to understand the population genetics of this species so that insights may be provided on dispersal ability and connectivity for management and conservation purposes. The main aim of this study was therefore to identify the pattern of spatial and temporal variation along the west coast of South Africa with the main research questions being: How many populations can be identified along the coast of South Africa in the hottentot seabream? What is the spatial genetic variation within this population/s? What is the temporal genetic variation between years? In this study, I present the first population genetics study on the hottentot seabream (*Pachymetopon blochii*) using 14 highly polymorphic microsatellites (see Chapter 2) and several locations along the west coast of South Africa. I identify high levels of gene flow with weak spatial and temporal variation and hypothesize that the combination of life-history traits of this species (multiple spawning, pelagic larvae, limited adult movement and longevity) and oceanographic features ("jet" current transporting eggs and larvae) contributes to extensive mixing along the coast of South Africa.

3.2 Materials and Methods

3.2.1 Sampling

Samples were collected for the hottentot seabream during their summer peak spawning period (October to February). In both sampling years all samples were caught in November. A total of 288 samples were collected in 2001 from nine locations and 189 samples from six locations in 2009 (Fig. 3.1; Table 3.1). Four locations were sampled in both 2001 and 2009 to test for temporal stability (Table 3.1). All fish included in this study were larger than 20 cm fork length (FL) and were therefore considered as sexually mature adults. Fin clips were taken from each individual and the FL was recorded. The fin clips were then immediately stored in 95 % ethanol and subsequently stored at -20°C.

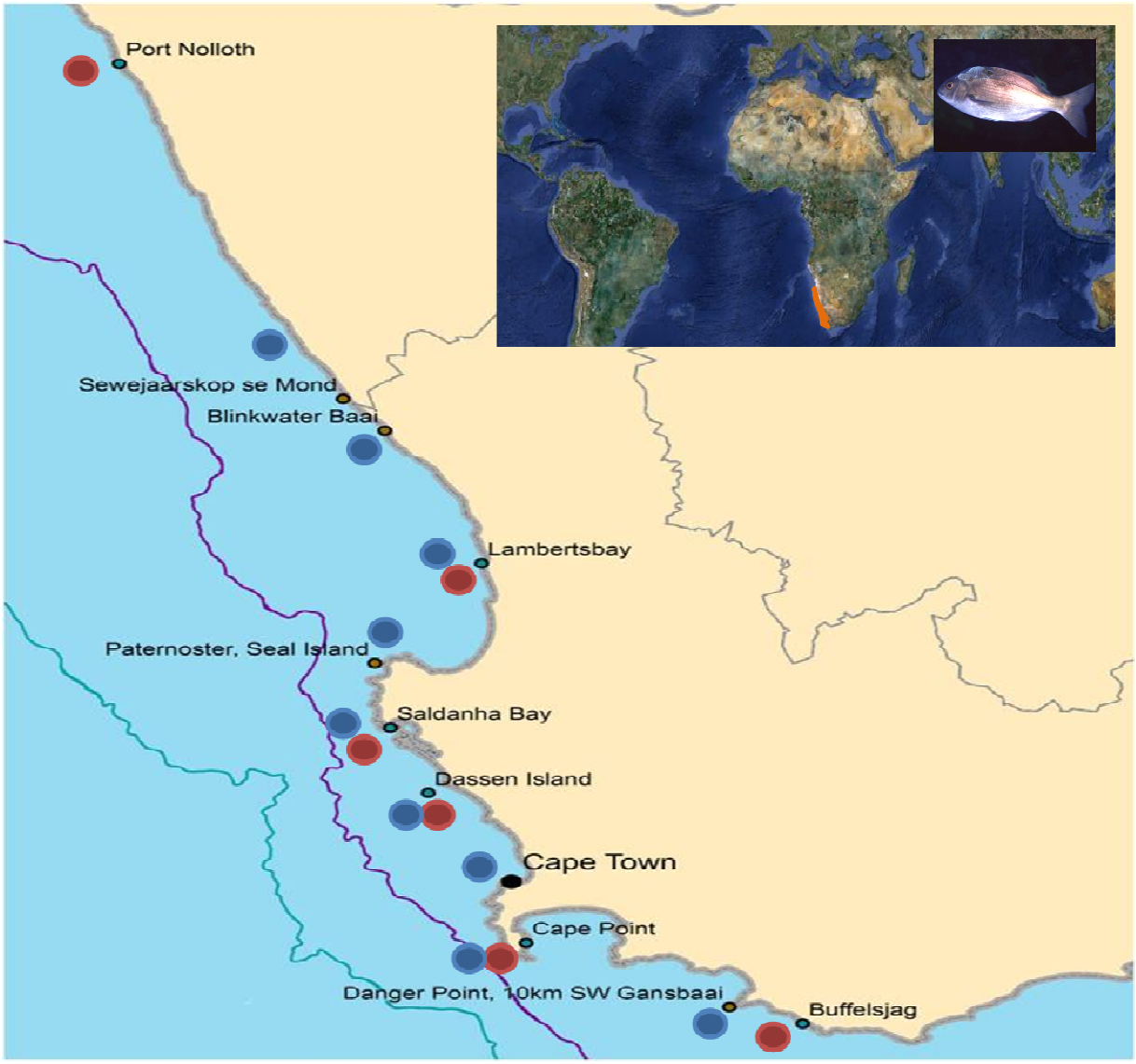


Fig. 3.1 Map of the locations sampled along the coast of South Africa in 2001 and 2009. The 2001 sampling locations are indicated in blue and the 2009 sampling locations in red. The inset shows a map of Africa (google maps; orange line indicates the west coast of southern Africa where the hottentot is distributed) and an image of the hottentot seabream.

3.2.2 Molecular markers

Seven microsatellite markers were developed in this species (PBt001, PBt003, PBt007, PBt011, PBt013, PBt014 and PBt018) and seven markers were identified through cross-species amplification (LLtr004, LLt005, LLt006, PB106, PA535, CL011 and SA043; Chapter 2). Only tetranucleotide and trinucleotide markers were used to reduce microsatellite genotype scoring errors through slippage during the Polymerase Chain Reactions (PCRs).

Table 3.1: A summary of the sampling locations, geographic co-ordinates, year sampled and number of individuals sampled (also see Fig. 3.1).

Geographic location	Locality identification	Longitude	Latitude	Year	No of individuals
Sewejaarskop	SJ1	17.7307	-31.1303	2001	29
Blinkwater Baai	BW1	17.8697	-31.2838	2001	22
Lamberts Bay	LB1	18.2715	-31.9320	2001	45
Paternoster	PA1	17.8666	-32.7433	2001	32
Saldanha Bay	SB1	17.9292	-33.0457	2001	32
Dassen Island	DI1	18.0597	-33.4287	2001	32
Cape Town	CT1	18.3505	-33.9783	2001	32
Cape Point	CP1	18.4808	-34.3367	2001	32
Danger Point	DP1	19.3030	-34.6300	2001	32
Port Nolloth	PN2	16.5144	-29.1515	2009	30
Lamberts Bay	LB2	18.1749	-32.0608	2009	33
Saldanha Bay	SB2	17.5610	-33.0224	2009	30
Dassen Island	DI2	18.0508	-33.2440	2009	30
Cape Point	CP2	18.3348	-34.2258	2009	36
Buffelsjag	BJ2	19.3408	-34.4335	2009	30

3.2.3 DNA extraction and microsatellite amplification

All DNA was extracted from fin clips using the Qiagen DNeasy extraction kit according to the manufacturer's specifications (Qiagen). The markers were labelled and designed into three multiplexes indicated in Appendix 3A and amplified using the Quantitect multiplex kit (Qiagen). The PCRs were amplified according to the manufacturer's specifications (Qiagen). Ten micro litre (µl) reaction volumes were used to amplify multiple microsatellite markers in a single reaction including 20 to 50 ng of genomic DNA, half the final reaction volume of Quantitect multiplex master mix, 0.2 pmol forward and 0.2 pmol reverse primer for each primer set. The amplification protocol included 95 ℃ for 15 minutes and 45 cycles of 94 ℃ for 60 s and 60 ℃ for 90 s. The PCR products were gel electrophoresed on 2 % agarose gels (Separations) and visualized with GelRed™ Acid stain (Biotium). The products were analysed on an ABI 3100 automated sequencer (Applied Biosystems, Foster City, CA) with GeneScan Liz™ 500 Size Standard (Applied Biosystems). Genotypes were scored using GENEMARKER version 1.5 software (SoftGenetics, State College, Pennsylvania, USA). A set of 40 randomly chosen individuals were re-genotyped with all markers to assess genotyping error as indicated in chapter 2.

3.2.4 Summary statistics

Individuals that were lacking more than two monolocus genotypes were discarded from the data set to reduce the amount of missing data. The summary statistics for each location sampled and cohorts were: Number of alleles (N_A), allelic richness (A_R), observed heterozygosity (H_O), expected heterozygosity (H_E), inbreeding coefficient (F_{IS}) and pairwise genetic differentiation (F_{ST}; Raufaste & Bonhomme 2000) and deviation from Hardy-Weinberg Equilibrium (HWE). These parameters were calculated with GENETIX version 4.05 software (Belkhir *et al.* 1996-2004). Genetic differentiation between populations estimated by F_{ST} was according to Raufaste and Bonhomme (2000) which is more accurate than F_{ST} by Weir & Cockerham (1984) when working with high gene flow species such as marine fish (Waples & Gaggiotti 2006). Other coefficients and tests such as R_{ST} and exact test were determined with ARLEQUIN version 3.11 (Excoffier *et al.* 2005). Overall F_{ST}, F_{IS} and F_{IT} were determined in GENETIX with 10 000 permutations according to Weir & Cockerham (1984). Deviations from HWE were calculated with GENEPOP version 4.0.10 (Rousset 2008), which determines the *f* of deviation from HWE following Weir and Cockerham (1984). The parameters were 1000 dememorizations, 100 batches, with 1000 iterations per batch. The incidence of null alleles was tested using MICRO-CHECKER version 2.2.3 (van Oosterhout *et al.* 2004). Linkage disequilibrium between individuals and markers was tested in GENEPOP using 10 000 steps and 10 000 iterations. When multiple statistical tests were performed Bonferroni correction was implemented (Rice 1989).

3.2.5 Inferring the number of populations

The program STRUCTURE version 2.3.2 (Pritchard *et al.* 2000), which is a Bayesian clustering analysis method that estimates the most likely number of populations, was used to infer the number of populations of the hottentot seabream. The analysis was done with an admixture model and correlated allele frequency model. It was then also performed with sampling locations as prior (Hubisz *et al.* 2009). Populations (*K*) were tested from 1 - 5 with a burn-in length of 100 000 and a MCMC of 100 000 steps then followed. The 20 iterations were performed for both 2001 and 2009 samples separately and the years combined as one population. The signal strength of the posterior probability was determined by the *ad hoc* statistic Δ*K* (Evanno *et al.* 2005); Δ*K* shows the rate of change of the log probability between the different assumed number of populations, the highest value (largest change of Δ*K*) indicates the most likely number of populations (Evanno *et al.* 2005).

3.2.6 Analysis of Molecular Variance (AMOVA)

Analysis of Molecular Variance (AMOVA; Excoffier *et al.* 1992) was calculated in ARLEQUIN and 10 000 permutations were used to investigate the different variance components tested. Proximate sampling locations were grouped: The groupings for 2001 were defined by pooling locations less than 300 km apart: (SJ1, BWB), (LB1, PA1), (SB1, DI1, CT1), (CP1, DI1) into four groups. The 2009 proximate groupings (less 300 km apart) were as follows (PN2), (LB2), (SB2, DI2) (CP2, BJ2). The locations were then grouped into North vs. South sampling locations for each year: The 2001 grouping was: (SJ1, BW1, LB1, PA1) and (SB1, DI1, CT1, CP1, DI1) and the 2009 grouping was (PN2, LB2) and (SB2, DI2, CP2, BJ2).

3.2.7 Factorial Correspondence Analysis (FCA)

Factorial Correspondence Analysis (FCA) was conducted in GENETIX based on the pairwise genetic distance between individuals. The analysis expresses the difference between individuals as factors and this explains the correspondence of individuals in three dimensions (3D). The inertia of locations sampled was used as an additional factor. In species with high gene flow, spatial information is valuable to identify any weak underlying structure. This analysis allows for a visual representation of the relationship between individuals and how they cluster together within a three-dimensional space.

The FCA was done for each year with each location considered a separate population, the data sets were combined and each year was considered a separate population. The individuals were then grouped into year classes according to the FL. Individuals were aged according to the growth curve determined through an otolith study by Pulfrich and Griffiths (1988b) with corresponding FL (Table 3.2). The age groups were 4 - 5, 5 - 6, 6 - 7, 7 - 8, 8 - 9, 9 - 10 and 10 – 11 years and were grouped into different age classes. The size and age classes were non-overlapping.

3.2.8 Isolation by Distance (IBD)

Isolation by distance was tested independently for each year sampled to avoid biases linked to the temporal survey. The geographic distance was determined from the distance along the coast line between sampling locations according to the co-ordinates. The measure of $F_{ST}/(1-F_{ST}$; Rousset 1997) was used as genetic distance and calculated in GENEPOP (Rousset 2008). Isolation by distance was tested using the Isolde option with a 1000 random permutations. This option tests the correlation using a Mantel test between the geographic distance (in kilometres between locations) and the genetic distance ($F_{ST}/(1-F_{ST})$) matrix.

Table 3.2 "Standard errors of the von Bertalanffy growth parameters and the lengths at age, calculated by the jack-knife method of variance estimation" taken from Pulfrich and Griffiths (1988b)

Age	No of individuals per age class 2001	2009	Jack-knife estimate Fork length (mm)	Standard error	Coefficient of variation
1	-	-	69.601	1.568	0.023
2	-	-	112.844	0.749	0.007
3	-	-	152.093	0.473	0.003
4	-	-	187.718	0.523	0.003
5	60	21	220.054	0.548	0.003
6	85	56	249.403	0.574	0.002
7	60	37	276.043	0.793	0.003
8	25	14	300.223	1.252	0.004
9	26	14	322.169	1.870	0.006
10	19	11	342.090	2.588	0.008
11	5	5	360.171	3.387	0.009
12	3*	2*	376.582	4.232	0.011

* These samples were not included in cohort analyses

3.2.9 Spatial Autocorrelation (SAC)

The spatial genetic structure was analysed using Spatial Autocorrelation (SAC; Smouse & Peakall 1999) available in GENALEX version 6 (Peakall & Smouse 2006). The spatial autocorrelation coefficient denoted as "r" was calculated using the genetic and geographic distance between individuals within distance classes (categories of geographic distance). A 95 % confidence interval was generated using 1000 random permutations and when the data deviate significantly from the null hypothesis (a random geographic distribution) a positive SAC is identified. Under the isolation by distance model, geographically proximate samples will have a positive SAC, which decreases as the geographic distances increases. The point where SAC is not detected (where "r" = 0), is known as the "patch size" or dispersal ability.

3.2.10 Effective population size (N_e)

The effective population size (N_e) was determined using the program MLNE which uses a pseudo-likelihood approach (Wang & Whitlock 2003). MLNE requires two temporally spaced samples, which provides information on the temporal variance in allele frequency. MLNE considered the data as one isolated population sampled temporally at two time periods, 2001 and 2009 respectively with no migration between populations. Therefore all location data was combined for a specific sampling year. The upper limit was chosen as 10 000 and tested a second time with an upper limit of 25 000. The number of generations between the two sampling years was determined as 1.5. This estimate was obtained assuming that the majority of individuals were spawning at about six years of age (Pulfrich & Griffiths 1988b), with an age of sexual maturity is at five years and a maximum age of twelve years.

3.3 Results

3.3.1 Summary statistics

No discrepancies were identified between the 40 individuals that were re-genotyped, suggesting that the data set is not affected by scoring errors or other artefacts causing genotyping errors. MICRO-CHECKER identified that all markers except LLt005 (19.4 %) showed negligible levels of null alleles (less than 7 % per locus) and therefore only LLt005 was removed from further analyses. The association between loci (linkage disequilibrium) based on their alleles were tested and 13 of 1132 tests showed a significant value (*P*-value < 0.05). There was no specific pattern of linkage disequilibrium and marker pairs did not show any significant linkage, therefore the remaining 13 loci were included in the subsequent analyses.

The F_{ST} and R_{ST} between locations and years are summarised in Table 3.3. Bonferroni corrections has been shown to be too stringent in species with a weak genetic structure, thus in Table 3.3 all significant F_{ST} and R_{ST} values are indicated (Moran 2003). The pairwise F_{ST} and R_{ST} values between and within the 2001 and 2009 sampling locations were very low (Table 3.3). LB2 showed significant R_{ST} values (*P*-value < 0.01) to SB2 and CP2, which are both more southerly locations along the west coast of South Africa. After Bonferroni correction only one pairwise R_{ST} was significant, namely between Cape Town (CT1) and Lamberts Bay (LB2) and one pairwise F_{ST} was significant between LB2 and Saldanha Bay (SB1). A pairwise F_{ST} between pooled years was 0.0017 and not significant.

Table 3.3 Pairwise F_{ST} (Raufaste and Bonhomme 2000) and R_{ST} values between locations and years; F_{ST} values are above and R_{ST} values are below the diagonal. Blue shading indicates the pairwise comparisons between 2001 locations, purple indicates pairwise comparisons between 2009 locations and pink cells are comparisons between years. Significant pairwise values are indicated in bold.

	SJ1	BW1	LB1	PA1	SB1	DI1	CT1	CP1	DP1	PN2	LB2	SB2	DI2	CP2	BJ2
SJ1	*	**0.023***	0.006	0.009	**0.017***	0.004	0.007	0.001	0.006	0.005	**0.030****	0.006	0.004	0.006	**0.018***
BW1	-0.008	*	0.004	**0.018***	0.016	0.019	0.013	0.008	0.008	**0.023***	0.017	0.007	0.017	0.005	0.005
LB1	0.005	-0.009	*	0.009	0.003	0.003	0.003	0.005	0.002	**0.015***	0.009	0.006	0.006	0.006	0.006
PA1	-0.006	-0.008	-0.001	*	0.013	0.009	0.008	0.012	**0.018***	0.013	0.006	**0.018***	0.009	**0.016***	0.014
SB1	-0.008	-0.018	-0.003	-0.002	*	0.003	**0.015***	0.001	0.009	**0.021****	**0.039*****	**0.017***	0.007	0.008	0.009
DI1	-0.009	-0.011	-0.001	-0.006	-0.010	*	0.007	0.003	0.005	0.007	0.009	0.006	0.007	0.007	0.013
CT1	0.014	0.015	0.003	0.002	**0.015***	0.004	*	0.004	0.005	0.010	0.017	0.008	0.003	0.007	0.009
CP1	-0.007	-0.008	0.001	-0.004	-0.005	-0.007	0.004	*	0.005	0.003	0.008	0.004	0.007	0.009	0.005
DP1	0.003	0.001	-0.010	-0.001	-0.001	-0.009	-0.003	-0.001	*	0.002	**0.032****	0.006	0.009	0.008	0.007
PN2	-0.003	-0.016	-0.010	-0.005	-0.011	-0.010	0.004	-0.007	-0.011	*	0.016	0.011	0.012	0.008	**0.023****
LB2	0.009	-0.007	0.007	0.014	0.003	0.013	**0.036****	0.002	0.016	-0.003	*	**0.034*****	0.012	**0.034****	0.008
SB2	**0.018***	-0.011	-0.001	**0.020***	0.000	0.013	**0.027***	**0.021***	0.008	-0.002	0.013	*	0.005	0.016	0.009
DI2	0.008	-0.017	-0.006	0.004	-0.008	0.001	0.009	-0.003	-0.002	-0.011	0.000	-0.005	*	0.014	0.009
CP2	0.016	-0.011	0.005	0.011	**0.017***	0.017	**0.032****	0.015	0.006	-0.004	0.012	**0.020***	0.014	*	0.01
BJ2	0.000	-0.014	-0.001	0.010	-0.012	-0.002	**0.023***	0.007	0.002	-0.008	0.007	-0.010	-0.006	0.008	*

*P-value < 0.05, **P-value < 0.01 and ***P-value < 0.001

The overall F_{ST}, F_{IS} and F_{IT} for the 2001 samples were -0.0002 (ns), 0.0368 (*P*-value < 0.001) and 0.0367 (*P*-value < 0.001) respectively. The 2009 samples had an overall F_{ST}, F_{IS} and F_{IT} of 0.001 (ns), 0.0229 (*P*-value < 0.01) and 0.0239 (*P*-value < 0.001) respectively.

A summary of the total number of alleles, allelic richness (A_R), average H_O, H_E and F_{IS} for all locations is indicated in Table 3.4. The average expected heterozygosity (H_E) ranged from 0.85 - 0.879 for the 2001 and from 0.844 - 0.877 for the 2009 locations. The values of F_{IS} ranged from 0.009 - 0.066 and -0.01 - 0.52 in 2001 and 2009 respectively with some deviations from HWE (*P*-value < 0.05). Summary statistics were determined for each marker and each location and is summarised in Appendix 3A. The H_E, H_O, A_R and F_{IS} was also determined for cohorts of both sampling years and then compared to the overall values for these parameters to determine if there is genetic variation within year classes (Table 3.5).

Table 3.4 The total number of alleles per location, A_R and the average H_O, H_E and F_{IS} over all markers. Deviations from HWE are indicated in bold.

	N	Na (total)	A_R	H_O (ave)	H_E (ave)	F_{IS} (ave)
SJ1	28	182	14	0.848	0.871	**0.044****
BW1	22	154	12	0.825	0.851	**0.052***
LB1	44	189	14.53	0.829	0.878	**0.066*****
PA1	32	183	14.07	0.85	0.869	**0.038****
SB1	31	178	13.69	0.865	0.878	**0.030***
DI1	32	183	14.1	0.884	0.879	0.009
CT1	32	175	13.46	0.845	0.865	**0.038***
CP1	32	184	13.38	0.859	0.872	**0.030***
DP1	30	188	14.6	0.892	0.879	0.01
PN2	28	174	13.38	0.878	0.877	0.019
LB2	19	140	10.76	0.83	0.844	**0.052***
SB2	27	172	13.15	0.89	0.866	-0.01
DI2	28	171	13.15	0.857	0.866	0.025
CP2	30	183	14.15	0.872	0.875	0.02
BJ2	28	164	12.61	0.857	0.864	**0.026***

**P*-value < 0.05, ***P*-value < 0.01 and ****P*-value < 0.001

Table 3.5 Summary of the H_O, H_E, A_R and F_{IS} over all markers for each size class in each year. Deviations from HWE are indicated in bold.

		2001					2009				
Size class	**Age (years)**	**N**	**H_E**	**H_O**	**A_R**	**F_{IS}**	**N**	**H_E**	**H_O**	**A_R**	**F_{IS}**
1	4-5	60	0.884	0.861	16.1	**0.034****	21	0.863	0.838	12	**0.054****
2	5-6	85	0.881	0.870	17.2	**0.018***	56	0.873	0.854	15.3	**0.030***
3	6-7	60	0.882	0.858	15.9	**0.035****	37	0.872	0.863	13.3	0.024
4	7-8	25	0.860	0.819	12.7	**0.067*****	14	0.867	0.900	10.8	0.000
5	8-9	26	0.868	0.860	13.4	0.030	14	0.857	0.910	10.6	-0.024
6	9-10	19	0.854	0.792	11.6	**0.099*****	11	0.848	0.855	9.6	0.039
7	10-11	5	0.790	0.876	6.3	0.002	5	0.804	0.907	6.6	0.017
	All	280	0.886	0.855	20.6	**0.036*****	158	0.884	0.867	19.3	**0.023****

**P*-value < 0.05, ***P*-value < 0.01 and ****P*-value < 0.001

3.3.2. Inference of the number of populations

The number of populations along the west coast of South Africa was determined using STRUCTURE and a single population was identified. The data sets were run as separate years and also combined (data not shown) but this led to the same conclusion of one population. The likelihood values were the highest for one population and in both years this was the population number with the smallest standard deviation (Fig. 3.2). The *ad hoc* statistic Δ*K* (Evanno *et al.* 2005) also indicated that the most likely number of populations is one as the rate of change between numbers of populations was low.

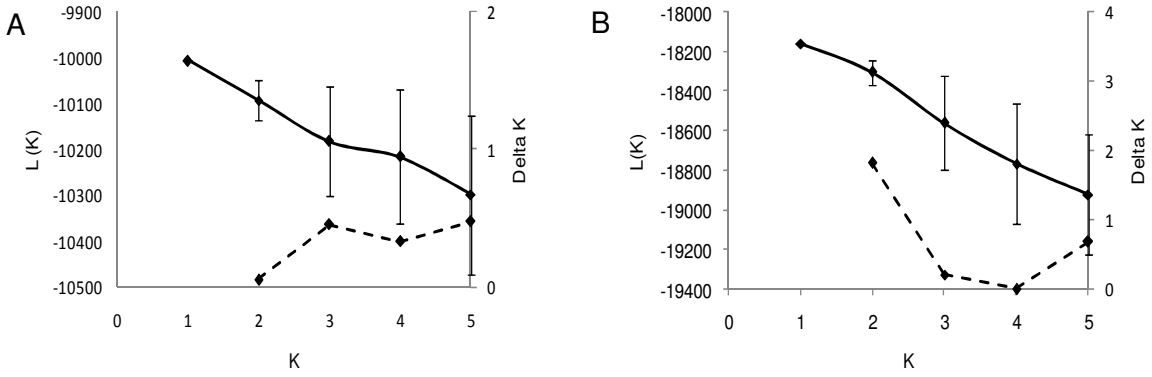


Fig. 3.2 Structure graphs for 2001 (**A**) and 2009 (**B**). In both graphs the left Y axis L (K) and the solid lines indicate the likelihood value of a specific number of populations (K) determined by STRUCTURE and the right Y axis and dashed trend lines indicate the *ad hoc* statistic Δ K.

3.3.3. Analysis of Molecular Variance

Spatial and temporal genetic variation was tested with AMOVA. In all cases the majority of the molecular variance was due to within location variation and not between locations or within groups (Table 3.6). However, even though low variance was identified between sampling locations (F_{ST}), both North vs. South grouping and proximate location groupings were all significant for the 2009 samples, indicating spatial structure. None of the groupings were significant for the 2001 samples.

Table 3.6 A summary of Analysis of Molecular Variance (AMOVA). The table indicates the source of genetic variation (% GV), the degrees of freedom (d.f.) and the fixation indices. Significant values are indicated in bold.

Source of variation	d.f.	% GV		Fixation indices
2009				
Between sampling locations	5	0.27	F_{ST}	**0.00270***
Within locations	314	99.73		
North vs. South grouping	1	0.15	F_{ST}	**0.00345***
Among locations within groups	4	0.19	F_{SC}	0.00194
Within locations	314	99.66	F_{CT}	0.00151
Proximate location groups	3	0.19	F_{ST}	**0.00299***
Among locations within groups	2	0.11	F_{SC}	0.00105
Within locations	314	99.70	F_{CT}	0.00193
2001				
Between sampling locations	8	0.01	F_{ST}	0.00007
Within sampling locations	557	99.99		
North vs. South groups	1	0.02	F_{ST}	0.00015
Among locations within groups	7	-0.00	F_{SC}	-0.00002
Within locations	557	99.99	F_{CT}	0.00017
Proximate location groups	3	0.01	F_{ST}	0.00009
Among locations within groups	5	0.00	F_{SC}	0.00000
Within locations	557	99.99	F_{CT}	0.00009

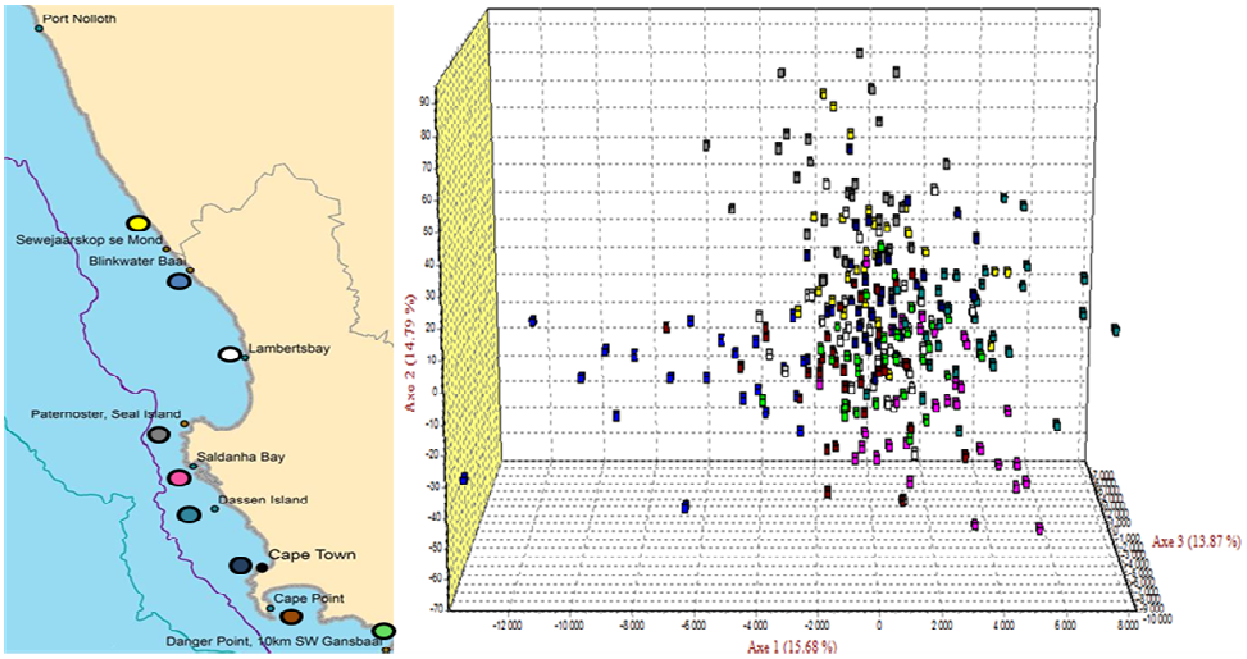
**P*-value < 0.05

3.3.4. Factorial Correspondence Analysis

The FCA between locations of each year indicated that there is no specific spatial variation among the 2001 samples (Fig. 3.3A). The 2009 locations show more spatial grouping with Port Nolloth (the most northerly location) and Lamberts Bay individuals clustering away from the main cluster of the more southerly locations (Fig. 3.3B). When all the samples were pooled according to the year (2001 and 2009), a shift is observed (Fig. 3.4A). The individuals were divided into year classes to identify any genetic variation between cohorts. Most of the individuals were in the 4 - 5, 5 - 6 or 6 - 7 year class, indicated by the yellow, blue and white

individuals (Fig. 3.4B). This showed that fewer individuals were caught of the older year classes in 2001. There was no variation within the different year classes for the 2001 samples (Fig. 3.4B). In 2009 the majority of the individuals were within the 5 - 6 and 6 - 7 year classes indicated by the blue and white respectively (Fig. 3.4C) and some separation is identified for the older year classes (seven to eleven years).

A



B

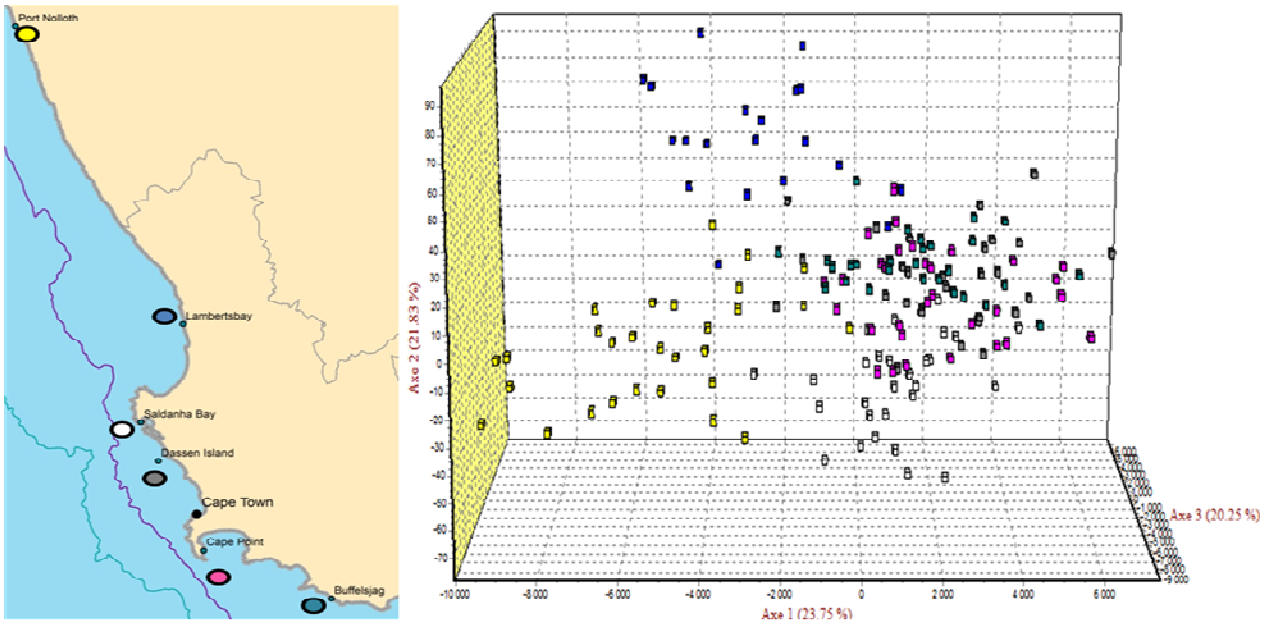


Fig. 3.3A and **B**: Factorial Correspondence Analysis between locations for A) 2001 locations and B) 2009 locations. Colours for locations on the map correspond to the colours in the FCA.

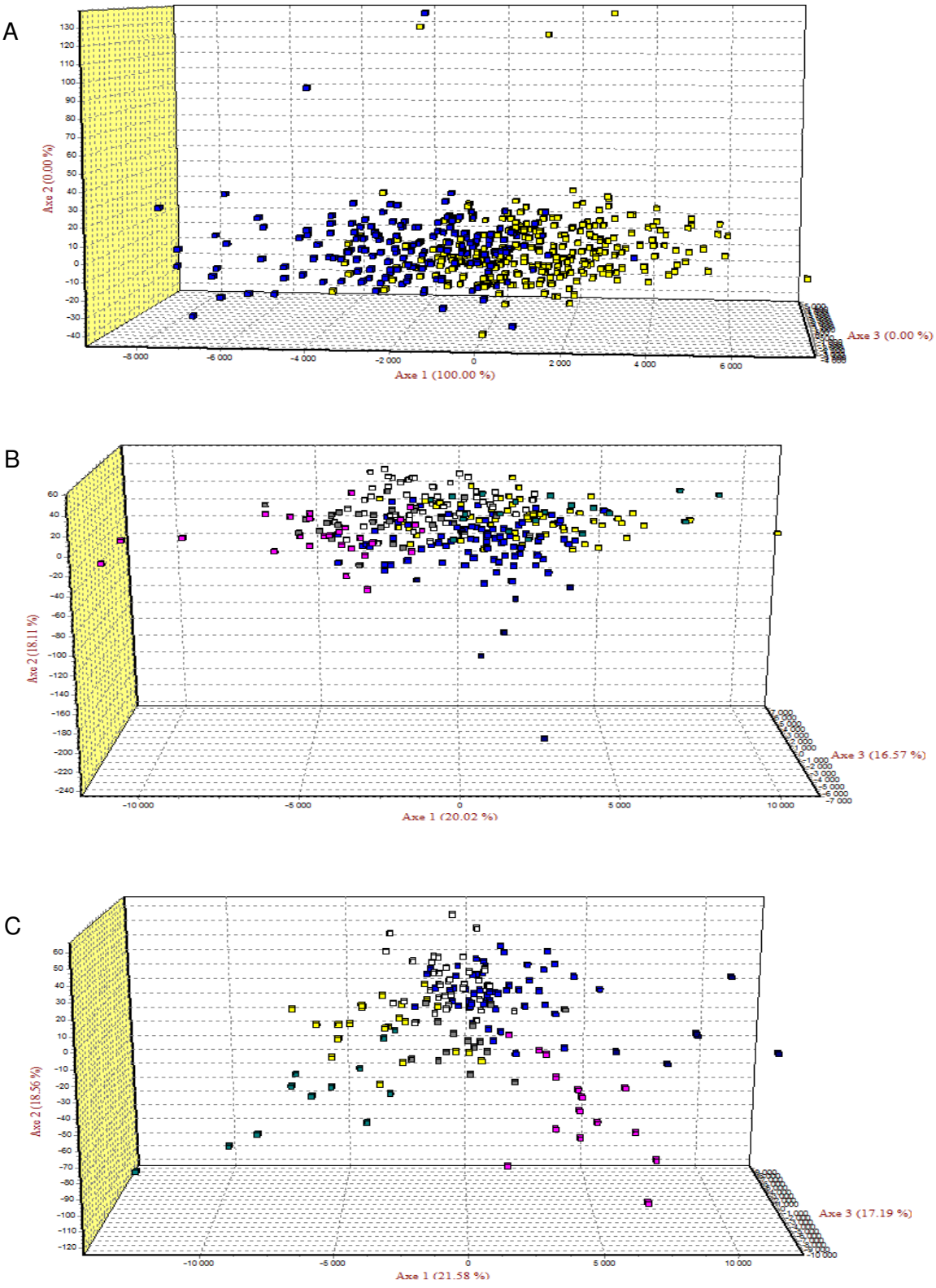


Fig. 3.4 A, **B** & **C**: Factorial Correspondence Analysis. **A**) 2001 (yellow) and 2009 (Blue) combined samples, **B**) 2001 samples and **C**) 2009 samples. In graphs B and C, yellow corresponds to 4 - 5 years, blue to 5 - 6 years, white to 6 - 7 years, grey 7 - 8 years, pink 8 - 9 years, turquoise 9 - 10 years and dark blue 10 - 11 years.

3.3.5. Isolation by distance

Sampling years were not combined but were plotted separately on the same axes. Neither year's samples showed a specific correlation between genetic and geographic distance along the west coast of South Africa. The null hypothesis could not be rejected which predicted that there was no correlation between geographic distance and genetic differentiation (*P*-values of 0.902 and 0.121 for 2001 and 2009 respectively).

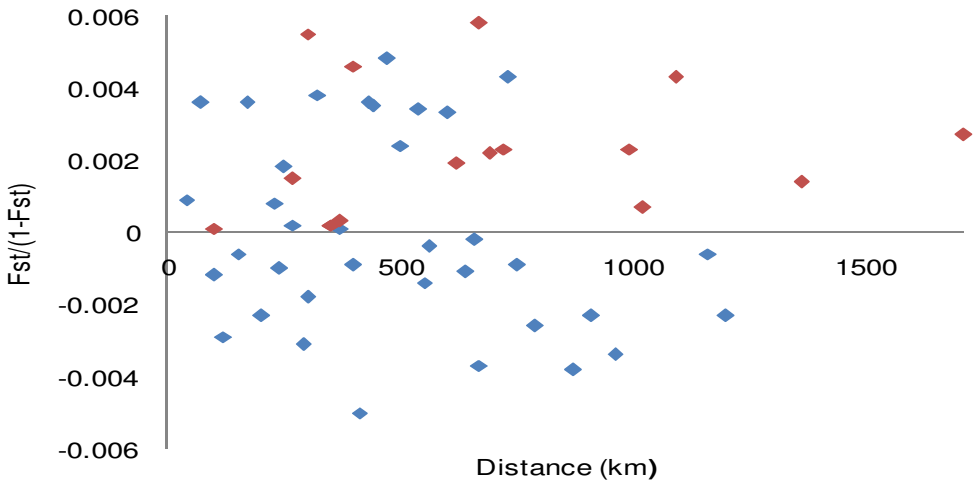


Fig. 3.5 Isolation by distance using linearized $F_{ST}/(1-F_{ST})$ and geographic distance. The correlation coefficients were 0.064 (*P*-value > 0.05) and 0.020 (*P*-value > 0.05) for 2001 and 2009 respectively. The blue points indicate the data for 2001 and the red for 2009.

3.3.6. Spatial Autocorrelation

Spatial autocorrelation was tested separately for the 2001 and 2009 years using specific distance classes determined by the distances between locations. These distances are summarised in Appendix 3B. Some markers were removed one at a time to indicate that no one specific marker was contributing to the observed pattern and the contributions of some individual markers are indicated in Appendix 3C for both 2001 and 2009 data sets.

In 2001 the spatial autocorrelation indicated that individuals that were 40 to 120 km (*P*-value = 0.005) and 730 - 900 km (*P*-value = 0.04) apart had a positive spatial autocorrelation (Fig. 3.6A). This indicates that this population does deviate slightly from the null hypothesis that assumes a random distribution of individuals (i.e. no spatial pattern) across the species range. In Fig. 3.6B all individuals from zero to 120 km and all individuals from zero to 230 km

apart showed a positive spatial autocorrelation (*P*-value = 0.017 and 0.026 respectively). The spatial auto-correlation coefficient "r" intercepts zero at about 300 km.

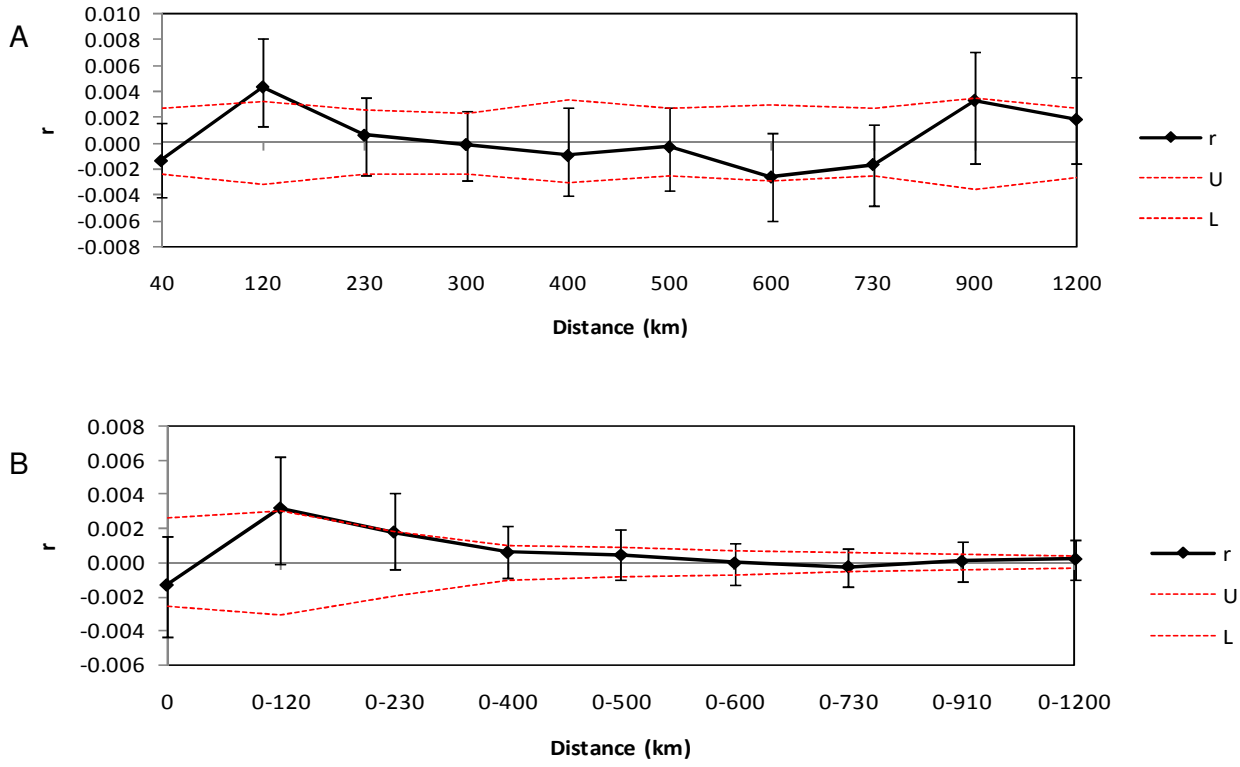


Fig. 3.6A and **B** Spatial autocorrelation graphs across size classes for 2001: **A**) Spatial autocorrelation of size classes determined by the distances between locations: **B**) All individuals included in each increased size class.

The 2009 samples had a positive spatial autocorrelation within locations (*P*-value = 0.005). However, across the distance classes there was no positive correlation (Fig. 7A). When all individuals were included from zero (*P*-value = 0.024) to 270 km (*P*-value = 0.007) there was a positive spatial autocorrelation and then it started to decrease (Fig. 7B). These analyses show that in 2009 there is some spatial structure and that this population significantly deviates from the null hypothesis. The "r" value intercepts zero at about 550 km. At 1710 km with all individuals of the study included there is a significant negative "r" value (*P*-value = 0.976), showing that individuals at these distances are less related than expected through random mating (i.e. there is population structure).

The overall spatial autocorrelation of both years does show deviations from the null hypothesis (i.e. no spatial structure). When all individuals were included from within locations (zero) to certain distance classes in both the 2001 and 2009 years, a "patch size" or predicted dispersal ability of larger than 300 km was indicated in both years (Fig. 3.6B and

3.7B). Also the patterns of spatial autocorrelation do not show the same pattern which indicates some temporal variation over the last eight years. It should be kept in mind that the 2009 sampling data covered a much larger part of the species' distribution range.

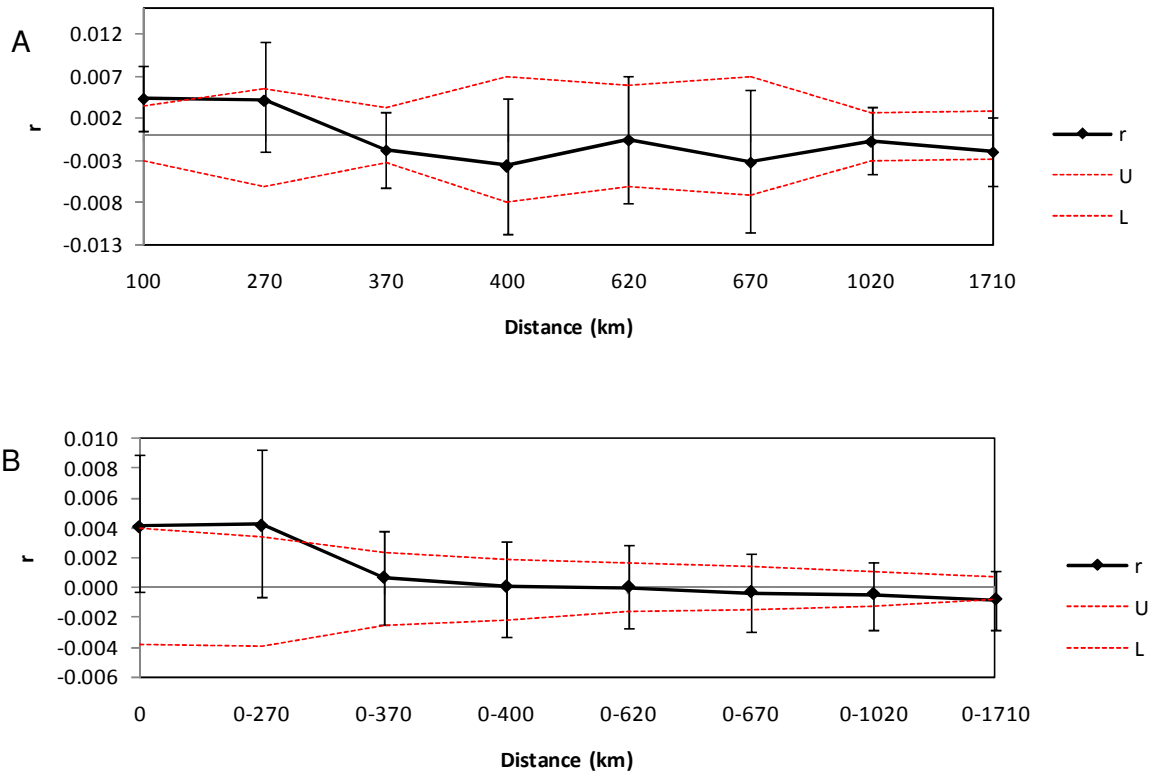


Fig. 3.7A and **B**: Spatial autocorrelation for the 2009 samples: **A**) Spatial autocorrelation of size classes determined by the distances between locations: **B**) All individuals included in each increased size class.

3.3.7. Effective population size

There is no known census population for the hottentot seabream but it has been recorded that approximately 600 tons are caught annually (Sauer *et al.* 2003), which is well over a hundred thousand fish per annum. The estimate of the effective population size (N_e) for the hottentot seabream (single population) was 9989 individuals with a confidence interval (CI) limit at 95 % of 10 000 individuals for the temporal sampling between 2001 and 2009 when the upper limit was set to 10 000. When the a priori upper limit was set to 25 000 the N_e was 24 990 and the CI was 25 000. This demonstrates that this specific data set is not sufficient to determine an accurate N_e for this species and these data will not be discussed further in the discussion.

3.4. Discussion

Hardy-Weinberg Equilibrium (HWE) predicts a panmictic population under the assumptions that reproduction is sexual with random mating, natural selection does not play a role, mutations are negligible, there is no migration and the population size is infinitely large (Halliburton 2004). In wild populations, true panmixia is unlikely to exist due to assumptions that cannot be met (infinite population size, random mating), so the first task of a population genetics study is to identify where the population structure of a species lies on the continuum from panmixia to isolated populations. Here, weak spatial and temporal structure was identified along the distribution of the hottentot seabream, allowing us to reject the hypothesis of a single well-mixed population (panmixia) along the west coast of South Africa.

3.4.1 Population identification

No more than one population could be definitively identified throughout the South African coastal distribution, although spatial and temporal variation was observed. Microsatellite markers provide information on a demographic scale and are known to have a lower power of resolution when dealing with large populations, as is found in many marine fish species (Waples & Gaggiotti 2006; Hauser & Carvalho 2008). This was supported by the STRUCTURE analysis (Fig 3.2A and B), overall F_{ST}, AMOVA (Table 3.5; 2001 data) and factorial correspondence analysis (FCA; Fig. 3.3A) which all indicated no clear separation of multiple populations. These results suggest high connectivity between the geographically separated coastal habitats of this relatively sedentary species, which is thought to be due to high fecundity, extended larval periods and oceanographic features.

Sheaves (2006) identified that in general across the sparid family the different species spawn when the sea temperatures are the lowest. This behaviour is believed to favour the survival of larvae, as the physiology of early life-history stages may have a low tolerance for warmer water (Sheaves 2006). This, in turn, can then be linked to dispersal ability as it has been shown that larvae adapted to colder water have longer larval durations. This is likely due to metabolic rates and development which decrease in colder water; the larvae may therefore take longer to develop physiologically (O'Connor *et al.* 2007) and these larvae will disperse further by ocean currents. The cold Benguela Current provides lower sea temperatures along the coast of South Africa which may contribute to an extended larval period. This, however, is probably not the only contributing factor as there are also a number of oceanographic features (currents, upwelling zones and eddies) which may enhance and restrict larval and adult movement.

Oceanographic features along the coast, such as the "jet" current are likely to transport pelagic eggs and larvae and deposit them in inshore habitats (refer to Fig. 1.2). This "jet" current is known to transport sardine and anchovy eggs and larvae from December to March (Hutchings *et al.* 2009) which coincides with the hottentot summer spawning period (Van der Elst 1988). Therefore we can conclude that it is likely that the hottentot pelagic eggs and larvae are being transported and distributed by the currents along the coast.

The hottentot seabream has life-history traits which contribute to many offspring, such as high fecundity (females may produce about a 100 000 eggs per spawning period) and multiple spawning periods. It is thought that the adults have limited dispersal ability (maximum dispersal distance recorded of 40 km after more than two years; Nepgen 1977), but due to the species' longevity, some connectivity may accumulate over the lifetime of individuals. Furthermore, the pelagic eggs and larvae transported and dispersed by specific oceanographic features may also contribute to extensive connectivity between more geographically isolated habitats along the coast.

3.4.2 Spatial genetic variation

This population exhibits weak spatial structure mainly for the 2009 samples. No genetic differentiation of populations was identified in 2001 as indicated by an overall F_{ST} of -0.0002 (*P*-value > 0.05). This was supported by the AMOVA where all the genetic variance was within locations rather than between locations and groupings (Table 3.6). The same result was suggested by the FCA, with no separation of locations (Fig. 3.3A) and a lack of significant isolation by distance. Spatial autocorrelation showed a deviation from panmixia and a dispersal ability of approximately 300 km (Fig. 3.6). This is the only evidence for deviation from panmixia in the 2001 data set.

The 2009 sampling showed a higher, though weak and non-significant spatial structure, with an overall F_{ST} of 0.001 (*P*-value > 0.05; Weir & Cockerham 1984). This was supported by a low but significant percent of variation between locations from AMOVA (equivalent to F_{ST}; Table 3.6; F_{ST} = 0.0027, *P*-value = 0.022). Moreover, the FCA showed a separation of Port Nolloth and Lamberts Bay from the more southerly locations and this was supported by a significant AMOVA when the North (PN2 & LB2) and South (SB2, DI2, CP, BJ) locations were grouped separately. There was no evidence of isolation by distance (Fig. 3.5). However, through spatial autocorrelation there was a significant deviation from the null hypothesis which predicts a random allocation of individuals along the west coast (Fig. 3.7).

Moreover, this method allowed for the estimation of patch size (dispersal ability), which was approximately 300 km, indicating spatial structure over a wide geographic distance.

Isolation by distance requires a strong signal across the entire geographic range whereas spatial autocorrelation is individual-based and works on both local and regional scales (Saillant *et al.* 2010). This makes spatial autocorrelation more sensitive to weak deviations from panmixia i.e. spatial structure. The "patch size" (potential dispersal ability) was estimated to be over 300 km in both years of sampling (Fig. 3.6B and 3.7B). This is further than what has been recorded in general across marine species where the mean distance varies from 10 - 150 km (Hauser & Carvalho 2008). However this is not applicable for all marine species as some may be able to disperse on a much wider geographical scale. The 2001 and 2009 spatial autocorrelation show similar patterns (Fig. 3.6 & 3.7) which supports a relatively stable population with a dispersal ability of about 300 km. This was also evident when the proximate 2009 locations were grouped (less than 300 km apart) and showed a significant amount of variation between groupings (Table 3.6).

The weak spatial structure identified for the 2009 samples which was not observed in the 2001 samples may be due to the wider distribution of sampled locations, some being more geographically isolated (Fig. 3.1). The Port Nolloth location was not sampled in 2001 and showed separation in the FCA (Fig 3.3B) from the rest of the sampling locations of 2009.

This weak structure only observed in 2009 may also be due to "random genetic patchiness" where, during different spawning years, due to variation in larval survival, diverse patterns of population structure may be observed that are not consistent over time and space (Selkoe *et al.* 2006) in the same geographic regions. This variation in reproductive success can be identified by determining the effective population size (N_e). However, the present data set could not be used to calculate an accurate N_e. Random genetic patchiness was tested by comparing general summary statistics between cohort groups and the overall sample sets (Table 3.5). This shows that in the 2001 and 2009 data sets there is more genetic diversity (seen in the allelic richness; A_R) in the overall population than in any of the cohorts (size-classes). Also for the 2009 data set only the youngest cohorts age (four to six) have significant F_{IS} values, as expected if there is variation in larval survival between years i.e. random genetic patchiness. This variation in reproductive success may lead to changes in the population structure over time and space and can possibly lead to genetic differentiation of adult populations (Pujolar *et al.* 2011).

There are two factors that are usually considered to influence the genetic patchiness, namely variation in reproductive success (where only a small portion of the adults spawn each year together with high larval mortality) and the inconsistent transport of these larvae caused by oceanographic features (environmental patchiness; Selkoe *et al.* 2006). Through the comparison of cohorts we have evidence for variation in reproductive success and known oceanographic features (jet current, Benguela current, upwelling zones) along the west coast (environmental patchiness) points to the potential change in structure we have observed between 2001 to 2009 being due to random genetic patchiness.

3.4.3 Temporal genetic variation

Only a weak signal of temporal variation was observed and this was between sampling years and among age classes sampled in 2009. No significant F_{ST} (0.0017, *P*-value > 0.05) between the sampling years (2001 and 2009) was observed. However, the FCA identified some spatial variation in 2009 that was not observed in 2001 (Fig. 3.3A and B). Within sampling years, when individuals were divided into year classes according to their fork length, there was no variation observed for 2001 (Fig. 3.4B) but there was some variation between the year classes sampled in 2009 (Fig. 3.4C). Explanations for these patterns of temporal variation may be that variations in reproductive success and oceanographic features over the last eight years have led to a change in the population structure. Within years, samples are caught from individuals that were spawned over a number of different years. As individuals get older there is less chance of them surviving thus changes in allele frequencies (random genetic drift) and more extensive mixing of adults occur (indicated by the non-significant F_{IS}, Table 3. 5).

3.4.4. Conclusion

In conclusion, weak spatial and temporal variation was identified along the west coast range of this species. The 2009 locations showed that individuals within locations were more closely related than between locations and also indicated that Port Nolloth may be relatively isolated. The results indicate that although there is weak genetic structure the population does not remain constant over time which suggests that variation in reproductive success together with environmental factors, such as changes in oceanographic features, influence the connectivity between habitats of this species. All these factors will influence the structure over time as well. Global warming may impact this species, as well as others occurring along the west coast of South Africa, as increasing sea temperatures may lead to shorter larval durations and more structured populations in the future.

3.5 References

Appleyard SA, Ward RD (2006) Genetic diversity and effective population size in mass selection lines of Pacific oyster (*Crassostrea gigas*). *Aquaculture* **254**, 148-159.

Arculeo M, Lo Brutto S, Sirna-Terranova M*, et al.* (2003) The stock genetic structure of two Sparidae species, *Diplodus vulgaris* and *Lithognathus mormyrus*, in the mediterranean sea. *Fisheries Research* **63**, 339-347.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996-2004) GENETIX 4.05 Logiciel sous Windows pour la Génétique des Populations, Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).

Burridge CP, Versace VL (2007) Population genetic structuring in *Acanthopagrus butcheri* (Pisces: Sparidae): Does low gene flow among estuaries apply to both sexes? *Marine Biotechnology* **9**, 33-44.

Chiba SN, Iwatsuki Y, Yoshino T, Hanzawa N (2009) Comprehensive phylogeny of the family Sparidae (Perciformes: Teleostei) inferred from mitochondrial gene analyses. *Genes Genet. Syst* **84**, 153-170.

De Innocentiis S, Lesti A, Livi S*, et al.* (2004) Microsatellite markers reveal population structure in gilthead seabream *Sparus auratus* from the Atlantic Ocean and Mediterranean Sea. *Fisheries Science* **70**, 852-859.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: as simulation study. *Molecular Ecology* **14**, 2611-2620.

Excoffier L, Laval G, Schneider S (2005) Arlequin version 3.11: An integrated software package for population genetics data analysis. *Evolution Bioinformatics* **1**, 47-50.

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

González-Wangüemert M, Cánovas F, Pérez-Ruzafa A, Marcos C, Alexandrino P (2010) Connectivity patterns inferred from the genetic structure of white seabream (*Diplodus sargus* L.). *Journal of Experimental Marine Biology and Ecology* **383**, 23-31.

González-Wangüemert M, Pérez-Ruzafa A, Cánovas F, Garcia-Charton JA, Marcos C (2007) Temporal genetic variation in populations of Diplodus sargus from the SW Mediterranean Sea. *Marine Ecology Progress Series* **334**, 237-244.

Gonzalez EG, Beerli P, Zardoya R (2008) Genetic structuring and migration patterns of Atlantic bigeye tuna, *Thunnus obesus* (Lowe, 1839). *BMC Evolutionary Biology* **252**, 1-14.

Halliburton R (2004) *Introduction to population genetics* Pearson Education, Inc. Pearson Prentice Hall.

Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries* **9**, 333-362.

Heemstra P, Heemstra E (2004) *Coastal Fishes of Southern Africa* South African Institute for Aquatic biodiversity and National Inquiry Service Centre.

Hogan JD, Thiessen RJ, Heath DD (2010) Variability in connectivity indicated by chaotic genetic patchiness within and among populations of a marine fish. *Marine Ecology Progress Series* **417**, 263-275.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9**, 1322-1332.

Hutchings L, van der Lingen CD, Shannon LJ*, et al.* (2009) The Benguela Current: An ecosystem of four components. *Progress in Oceanography* **83**, 15-32.

Kelly RP, Oliver TA, Sivasundar A, Palumbi SR (2010) A method for detecting population genetic structure in diverse, high gene-flow species. *Journal of Heredity* **101**, 423-436.

Moran MD (2003) Arguments for rejecting the sequential Bonferroni in ecological studies. *OIKOS* **100**, 403-405.

Nepgen CSDV (1977) The biology of the hottentot *Pachymetopon blochii* (Val.) and the silverfish *Argyrozona argyrozona* (Val.) along the Cape south-west coast. *Sea Fisheries Branch Investigational Report* **105**, 1-30.

Nielsen EE, Wright PJ, Hemmer-Hansen J*, et al.* (2009) Microgeographical population structure of cod *Gadus morhua* in the North Sea and west of Scotland: the role of sampling loci and individuals. *Marine Ecology Progress Series* **376**, 213-225.

O'Connor MI, Bruno JF, Gaines SD*, et al.* (2007) Temperature control of larval dispersal and the implications for marine ecology, evolution, and conservation. *PNAS* **104**, 1266-1271.

Patarnello T, Volckaert FAMJ, Castilho R (2007) Pillars of Hercules: is the Atlantic-Mediteranean transition a phylogeographic break? *Molecular Ecology* **16**, 4426-4444

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Planes S, Lemer S (2011) Individual-based analysis opens new insights into understanding population structure and animal behaviour. *Molecular Ecology* **20**, 187-189.

Planes S, Lenfant P (2002) Temporal change in the genetic structure between and within cohorts of a marine fish, *Diplodus sargus*, induced by a large variance in individual reproductive success. *Molecular Ecology* **11**, 1515-1524.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure using multilocus genotype data. *Genetics* **155**, 945-959.

Pulfrich A, Griffiths CL (1988a) The Fishery for Hottentot Pachymetopon blochii Val. in the South-Western Cape. *South African Journal of Marine Science* **7**, 227-241.

Pulfrich A, Griffiths CL (1988b) Growth, sexual maturity and reproduction in the hottentot *Pachymetopon blochii* (Val.). *South African Journal of Marine Science* **7**, 25-36.

Pujolar JM, Bevacqua D, Capoccioni F, *et al.* (2011) Genetic patchiness in European eel adults evidenced by molecular genetics and population dynamics modelling. *Molecular Phylogenetics and Evolution* **58**, 198-206

Punt AE, Pulfrich A, Butterworth DS, Penney AJ (1996) The effect of hook size on the size-specific selectivity of Hottentot *Pachymetopon blochii* (Val.) and on yield per recruit. *South African Journal of Marine Science* **17**, 155-172.

Raufaste N, Bonhomme F (2000) Properties of bias and variance of two multiallelic estimators of F_{ST}. *Theoretical Population Biology* **57**, 285-296.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution* **43**, 223-225.

Roberts DG, Ayre DJ (2010) Panmictic population structure in the migratory marine sparid *Acanthopagrus australis* despite its close association with estuaries. *Marine Ecology Progress Series* **412**, 223-230.

Rossi AR, Perrone E, Sola L (2006) Genetic structure of gilthead seabream, Sparus aurata, in the Central Mediterranean Sea. *Central European Journal of Biology* **1**, 636-647.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**, 1219-1228.

Rousset F (2008) Genepop'007: A complete reimplementation of Genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Saillant E, Bradfield SC, Gold JR (2010) Genetic variation and spatial autocorrelation among young-of-the-year red snapper (*Lutjanus campechanus*) in the northern Gulf of Mexico. *ICES Journal of Marine Science* **67**, 1240-1250.

Sauer WHH, Hetch T, Britz PJ, Mather D (2003) An economic and sectoral study of the South African fishing industry. *Fishery Profiles* **2**, 1-312.

Selkoe KA, Gaines SD, Caselle JE, Warner RR (2006) Current shifts and kin aggregation explain genetic patchiness in fish recruits. *Ecology* **87**, 3082-3094.

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**, 615-629.

Sheaves M (2006) Is the timing of spawning in sparid fishes a response to sea temperature regimes? *Coral Reefs* **25**, 655-669.

Shillington FA, Reason CJC, Duncombe Rae CM, Florenchie P, Penven P (2006) Large Scale Physical Variability of the Benguela Current Large Marine Ecosystems (BCLME). *Large Marine Ecosystems* **14**, 48-68.

Smouse P, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus allelic structure. *Heredity* **82**, 561-573.

Teske PR, Forget FRG, Cowley PD, von der Heyden S, Beheregaray LB (2010) Connectivity between marine reserves and exploited areas in the philopatric reef fish Chrysoblephus laticeps (Teleostei: Sparidae). *Marine Biology* **157**, 2029-2042.

Van der Elst R (1988) *A guide to the common sea fishes of Southern Africa*, 2nd edn. C. Struik Publishers.

van Oosterhout C, Hutchinson WF, Wills DP, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.

von der Heyden S, Lipinski MR, Matthee CA (2007) Mitochondrial DNA analyses of the Cape hakes reveal an expanding, panmictic population for *Merluccius capensis* and population structuring for mature fish in *Merluccius paradoxus*. *Molecular Phylogenetics and Evolution* **42**, 517-527.

Wang J, Whitlock MC (2003) Estimating Effective Population Size and Migration Rates from Genetic Samples Over Space and Time. *Genetics* **163**, 429-446.

Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity* **89**, 438-450.

Waples RS (2010) Spatial-temporal stratifications in natural populations and how they affect understanding and estimation of effective population size. *Molecular Ecology Resources* **10**, 785-796.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their connectivity. *Molecular Ecology* **15**, 1419-1439.

Ward RD, Woodwark M, Skibinski DOF (1994) A comparison of genetic diversity levels of marine, freshwater and anadromous fishes. *Journal of fish biology* **44**, 213-232.

White TA, Fotherby HA, Stephens PA, Hoelzel AR (2010) Genetic panmixia and demographic dependence across the North Atlantic in the deep-sea fish, blue hake (*Antimora rostrata*). *Heredity* **99**, 9999-9999.

3.6 Appendices

Appendix 3A Summary statistics for each location in 2001 and 2009. The table includes number of alleles (N_A), Observed heterozygosity (H_O), Expected heterozygosity (H_E) and F_{IS}. Deviations from HWE are also indicated.

Samples	*PBt001*[3]	*PBt003*[3]	*PBt007*[1]	*PBt011*[3]	*PBt013*[2]	*PBt014*[3]	*PBt018*[1]	*LLt006*[3]	*LLtr004*[1]	*PA535*[2]	*PB106*[1]	*SA043*[2]	*CLO11*[2]
No of samples	443	443	443	418	441	442	443	432	441	441	442	430	443
SJ1	**28**												
N_A	11	14	13	19	15	12	15	13	8	22	9	21	10
H_O	0.892	0.857	0.892	0.928	0.821	0.821	0.785	0.821	0.714	0.928	0.821	0.892	0.857
H_E	0.836	0.892	0.892	0.911	0.844	0.885	0.906	0.856	0.814	0.912	0.829	0.916	0.833
F_{IS}	-0.050	0.057	0.018	-0.001	0.045	0.091	0.151	0.059	0.141*	-0.000	0.028	0.044	-0.010
BW1	**22**												
N_A	10	14	8	15	14	10	15	10	8	16	9	16	9
H_O	0.818	0.954	0.863	0.818	0.727	0.863	0.818	0.647	0.863	0.727	0.909	0.90	0.818
H_E	0.824	0.889	0.815	0.882	0.815	0.858	0.891	0.863	0.793	0.88	0.807	0.917	0.836
F_{IS}	0.031	-0.050	-0.036	0.096	0.131	0.017	0.105	0.279	-0.065	0.205***	-0.102	0.032	0.045
LB1	**44**												
N_A	12	17	13	16	17	12	17	11	10	20	9	25	10
H_O	0.863	0.909	0.750	0.681	0.931	0.818	0.931	0.750	0.704	0.977	0.886	0.878	0.704
H_E	0.838	0.919	0.886	0.919	0.880	0.874	0.908	0.879	0.816	0.894	0.824	0.942	0.836
F_{IS}	-0.018	0.023	0.165	0.269***	-0.046	0.076	-0.014	0.159*	0.148	-0.081	-0.064	0.080*	0.169
PA1	**32**												
N_A	10	16	10	15	14	10	20	10	11	21	10	27	9
H_O	0.750	0.875	0.650	0.937	0.875	0.906	0.906	0.781	0.843	0.906	0.781	0.903	0.937
H_E	0.805	0.875	0.849	0.910	0.895	0.864	0.910	0.862	0.815	0.911	0.808	0.946	0.856
F_{IS}	0.084	0.016	0.243***	-0.014	0.039	-0.032	0.021	0.110	-0.019	0.021	0.050	0.063	-0.078
SB1	**31**												
N_A	10	14	13	16	16	12	16	12	9	20	8	22	10
H_O	0.838	0.838	0.935	0.871	0.871	0.871	0.871	0.774	0.838	0.967	0.806	1.00	0.774
H_E	0.827	0.899	0.860	0.905	0.911	0.864	0.906	0.879	0.846	0.929	0.803	0.943	0.845
F_{IS}	0.003	0.083	-0.071	0.055	0.061	0.009	0.055	0.136	0.025	-0.025	0.013	-0.044	0.101

DI1	**32**												
N_A	12	17	10	15	18	11	16	10	10	21	8	23	12
H_O	0.875	0.875	0.937	0.843	0.875	0.937	0.906	0.781	0.906	0.937	0.906	0.871	0.843
H_E	0.837	0.916	0.862	0.907	0.895	0.870	0.914	0.861	0.839	0.922	0.837	0.935	0.837
F_{IS}	-0.028	0.061	-0.071	0.086	0.038	-0.061	0.025	0.109*	-0.064	-0.001	-0.066	0.085	0.008
CT1	**32**												
N_A	11	16	12	15	14	10	17	12	10	18	8	22	10
H_O	0.750	0.906	0.843	0.937	0.781	0.875	0.875	0.656	0.781	0.875	0.968	0.937	0.812
H_E	0.774	0.912	0.838	0.902	0.844	0.864	0.904	0.877	0.856	0.897	0.830	0.934	0.823
F_{IS}	0.047	0.022	0.009	-0.023	0.091	0.004	0.049**	0.267*	0.103	0.040	-0.152	0.012	0.030
CP1	**32**												
N_A	11	17	10	12	17	12	19	16	8	21	9	22	10
H_O	0.843	0.906	0.937	0.906	0.906	0.906	0.968	0.687	0.781	0.906	0.843	0.871	0.718
H_E	0.792	0.920	0.862	0.900	0.874	0.865	0.928	0.873	0.826	0.911	0.821	0.931	0.838
F_{IS}	-0.050	0.031	-0.071	0.009	-0.021	-0.031	-0.028	0.228***	0.070	0.021	-0.011	0.089*	0.158*
DP1	**30**												
N_A	10	16	10	17	19	12	17	13	10	21	8	24	11
H_O	0.900	0.933	0.833	0.966	0.800	0.933	0.866	0.876	0.833	0.966	0.900	0.900	0.900
H_E	0.838	0.912	0.867	0.912	0.866	0.874	0.911	0.895	0.833	0.899	0.823	0.940	0.860
F_{IS}	-0.057	-0.006	0.056	-0.043	0.094	-0.050	0.066**	0.160*	0.018	-0.058	-0.076	0.059***	-0.030
PN2	**28**												
N_A	11	16	12	13	17	13	14	10	9	18	8	23	10
H_O	0.821	0.964	1.00	0.875	0.892	0.892	0.964	0.760	0.714	0.821	0.750	1.000	0.964
H_E	0.859	0.917	0.890	0.903	0.854	0.880	0.895	0.859	0.846	0.892	0.827	0.935	0.850
F_{IS}	0.063	-0.033	-0.104	0.053*	-0.027	0.004	-0.059	0.136	0.174*	0.098	0.112	-0.049	-0.116
LB2	**19**												
N_A	10	13	11	5	11	12	16	9	8	14	6	16	9
H_O	0.900	0.900	0.900	1.000	0.736	0.789	0.850	0.650	0.750	0.850	0.800	0.722	0.950
H_E	0.813	0.891	0.887	0.777	0.826	0.850	0.922	0.825	0.825	0.885	0.763	0.876	0.837
F_{IS}	-0.081	0.016	0.012	-0.091	0.136	0.098*	0.104*	0.236*	0.116	0.065	-0.022	0.204	-0.109
SB2	**27**												
N_A	14	16	12	14	15	11	15	11	9	16	11	18	10
H_O	0.963	0.814	0.925	0.888	0.888	0.888	0.888	0.814	0.888	0.888	0.851	0.884	1.00
H_E	0.864	0.879	0.866	0.895	0.846	0.851	0.902	0.867	0.804	0.862	0.846	0.920	0.856
F_{IS}	-0.096	0.092	-0.049	0.027	-0.031	-0.025	0.034	0.080	-0.086	-0.012	0.012	0.058	-0.150
DI2	**28**												
N_A	9	17	10	14	17	10	15	12	8	19	10	21	9
H_O	0.892	0.892	0.892	0.821	0.821	0.857	0.857	0.814	0.857	0.892	0.8214	0.807	0.928

H_E	0.821	0.914	0.843	0.906	0.868	0.841	0.899	0.851	0.801	0.906	0.853	0.922	0.836
F_{IS}	-0.069	0.042	-0.041	0.112*	0.072	-0.001	0.066	0.062	-0.052	0.034	0.056*	0.144*	-0.092
CP2	**30**												
N_A	10	14	12	17	16	10	21	12	7	19	10	23	12
H_O	0.833	0.866	0.900	0.958	0.933	0.900	0.866	0.800	0.666	0.928	0.896	0.866	0.933
H_E	0.829	0.891	0.857	0.898	0.893	0.862	0.924	0.868	0.820	0.913	0.855	0.935	0.838
F_{IS}	0.012	0.044	-0.032	-0.045	-0.028	-0.027	0.079	0.096	0.204***	0.002	-0.030	0.090*	-0.097
BJ2	**28**												
N_A	11	13	11	16	15	9	16	11	9	17	8	17	11
H_O	0.857	0.857	0.785	0.964	0.888	0.892	0.892	0.769	0.750	0.928	0.821	0.892	0.857
H_E	0.836	0.892	0.868	0.899	0.872	0.854	0.883	0.872	0.819	0.855	0.825	0.912	0.854
F_{IS}	-0.007	0.058	0.113	-0.054	0.000	-0.027	0.007	0.138*	0.103	-0.068	0.023	0.039	0.014

The numbers above the microsatellite markers indicate which multiplex they were in. **P*-value < 0.05, ***P*-value < 0.01 and ****P*-value < 0.001

Appendix 3B Distance matrix between 2001 localities and 2009 localities

2001

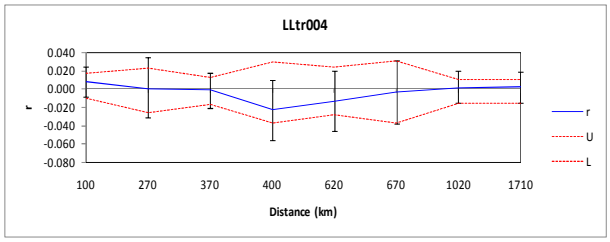
	SJKSM	BWB	LB	PAT	SB	DI	CT	CP	DP
SJKSM	0								
BWB	40	0							
LB	240	200	0						
PAT	470	430	230	0					
SB	540	500	300	70	0				
DI	640	600	400	170	100	0			
CT	790	750	550	320	250	150	0		
CP	910	870	670	440	370	270	120	0	
DP	1200	1160	960	730	660	560	410	290	0

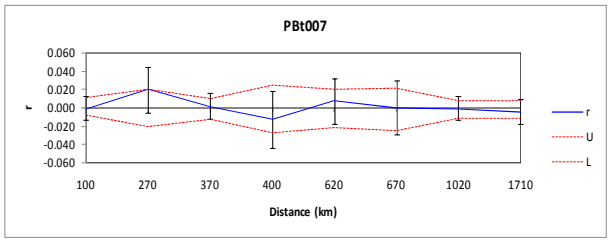
2009

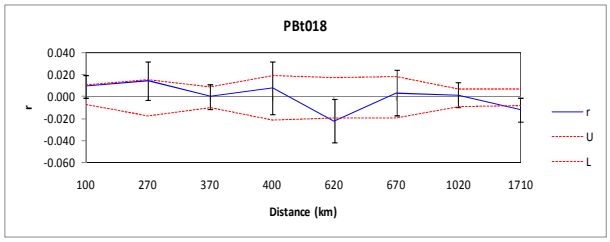
	PN	LB	SB	DI	CP	BJ
PN	0					
LB	690	0				
SB	990	300	0			
DI	1090	400	100	0		
CP	1360	670	370	270	0	
BJ	1710	1020	720	620	350	0

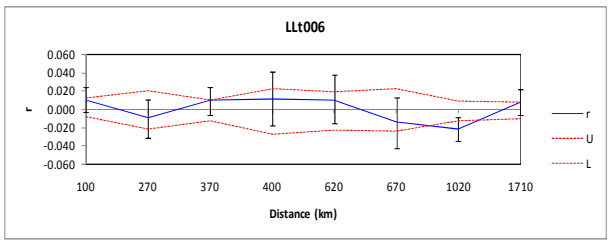
Appendix 3C Contributions of individual markers to the SAC (not all data is included) for each year.

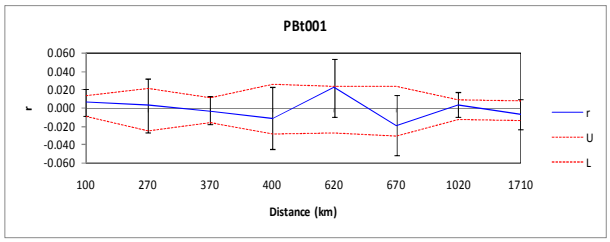
2009

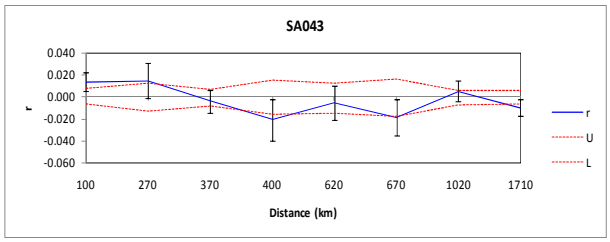




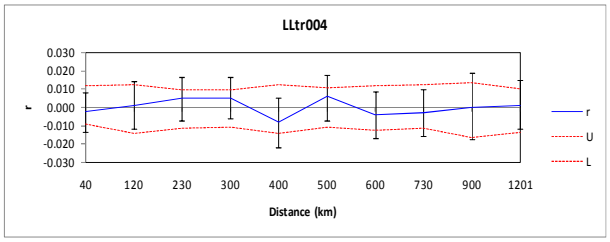


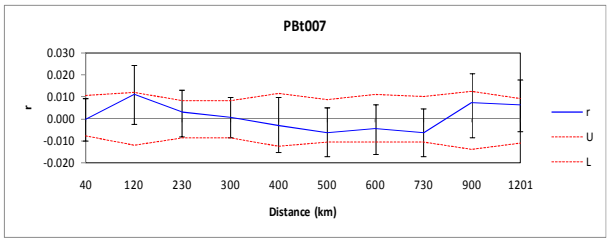


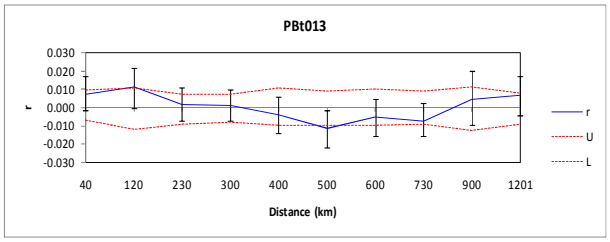


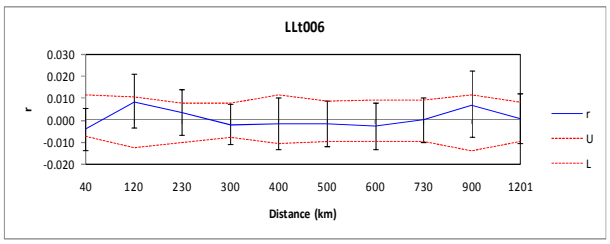


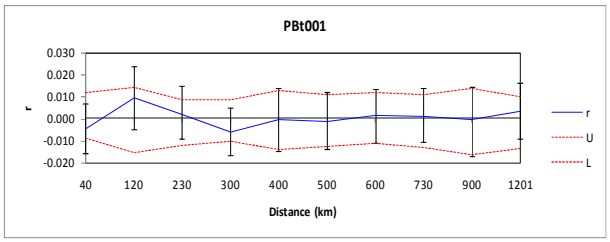
2001

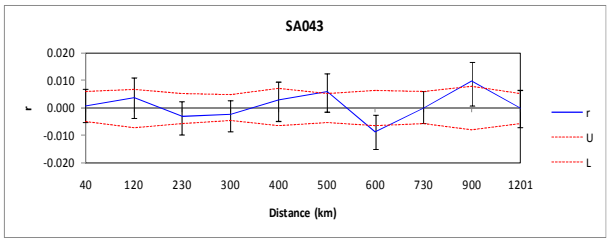












Chapter 4

Dissertation synthesis

Population genetics has contributed immensely to the understanding of gene flow and connectivity in the marine environment. Although it is difficult to monitor marine species in their natural habitats, molecular tools have provided a method for uncovering movement and interactions of marine populations (Liu & Cordes 2004). Before molecular analyses may be done an appropriate high-resolution set of molecular markers is required.

For this study I chose to develop microsatellite markers for their multitude of advantages such as co-dominance, high polymorphism and potential of cross-species amplification. Through the microsatellite development we were able to identify adequate sets of molecular markers (> 10 loci) for both the hottentot and white steenbras. Furthermore, a number of these markers worked across a wide range of economically important South African sparid species from both clade A and B (Chiba *et al.* 2009). As a result, these markers will not only benefit the specific species that the markers were developed in (hottentot and white steenbras) but will also be beneficial for a number of other sparids throughout the family with reduced costs and time of development.

Through this study I was also able to highlight a number of problems associated with microsatellites and molecular methods. By sequencing some of the alleles I could compensate for some of the major issues that affect microsatellite markers (e.g. homoplasy). Through the sequencing of a random set of alleles I confirmed that (1) the variation of the microsatellites was in most cases due to changes in the number of repeats, (2) that pyrosequencing may not always provide the most accurate sequences for primer design if not done at adequate depths of coverage, and (3) for some markers, the variations occurred within the flanking regions instead of the repeat unit. Future work should evaluate the real biases of using such markers when assessing weak population structure.

The aim of the second part of my study was to identify the spatial and temporal genetic variation in the hottentot seabream around South Africa using these microsatellite markers. All the results supported only weak spatial and temporal genetic variation for this species. This suggested that there is high connectivity across its South African distribution. Once life-history traits of this species and oceanographic features were considered it became clear that these factors most likely contribute to the mixing observed between localities. This species is the second most commonly caught line fish around the western Cape (Sauer *et al.*

2003) and the result obtained in our study helps to better understand its ecology and biology, which will be considered in further conservation and management strategies.

4.1 Future work

In terms of the markers identified through cross-species testing, a number of these markers are already being used in current South African population genetic studies. These species being studied are *Lithognathus lithognathus* (white steenbras), *Acanthopagrus berda* (riverbream), *Polysteganus undulosus* (seventyfour), *Diplodus cervinus hottentotus* (zebra seabream) and *Diplodus sargus capensis* (blacktail).

Only a limited amount of biological information was available for the hottentot and this first genetic study focusing on identifying population genetic structuring provides new perspectives. The high connectivity suggested by our results needs to be confirmed by (1) a local-scale temporal genetic survey together with (2) ecological surveys assessing the movement of pelagic eggs and larvae through inshore currents. Also a mitochondrial study should be considered to contribute to the nuclear marker data set as it will provide more information on historic events.

4.2 References

Chiba SN, Iwatsuki Y, Yoshino T, Hanzawa N (2009) Comprehensive phylogeny of the family Sparidae (Perciformes: Teleostei) inferred from mitochondrial gene analyses. *Genes Genet. Syst* **84**, 153-170.

Liu ZJ, Cordes JF (2004) DNA markers technologies and their applications in aquaculture genetics. *Aquaculture* **238**, 1-37.

Sauer WHH, Hetch T, Britz PJ, Mather D (2003) An economic and sectoral study of the South African fishing industry. *Fishery Profiles* **2**, 1-312.